PTO/PCT Rec'd 30 MAY 2000
PCT/US98/25314
09/555362

cDNAs AND PROTEINS INVOLVED IN HYPOXIA, CIRCADIAN AND ORPHAN SIGNAL TRANSDUCTION PATHWAYS, AND METHODS OF USE

Pursuant to 35 U.S.C. §202(c), it is acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Institutes of Health, Grant Nos. P30-CA07175 and ES05703.

This application claims priority to U.S. Provisional Application Serial No. 60/066,863, filed November 28, 1997, which is incorporated by reference herein.

FIELD OF THE INVENTION

This invention relates to the field of molecular signaling and physiological responses to external stimuli. In particular, this invention provides nucleic acid molecules and proteins that constitute new members of the bHLH-PAS superfamily of transcription regulators.

BACKGROUND OF THE INVENTION

Several publications are referenced in this application to describe the state of the art to which the invention pertains. Each of these publications is incorporated by reference herein.

The aryl hydrocarbon receptor (AH receptor or AHR), AH receptor nuclear transporter (ARNT), *Drosophila single-minded* gene product (SIM) and *Drosophila period* gene product (PER) are the founding members of an emerging superfamily of regulatory proteins. The AHR and ARNT are heterodimeric partners that transcriptionally upregulate genes involved in the metabolism of xenobiotics. The AHR is activatable by a number of widespread environmental pollutants like 2,3,7,8-

tetrachlorodibenzo-*p*-dioxin (TCDD). In the absence of agonist, the AHR is primarily cytosolic and functionally repressed, presumably as the result of its tight association with Hsp90. Current models suggest that agonist binding initiates translocation of the receptor complex to the nucleus and concomitantly weakens the AHR-Hsp90 association. Within the nucleus, Hsp90 is displaced and the AHR dimerizes with its partner ARNT resulting in a bHLH-PAS heterodimer with binding specificity for DNA sequences within enhancer elements upstream of gene products that metabolize foreign chemicals. In *Drosophila*, SIM is master regulator of midline cell lineage in the embryonic nervous system. *In vitro* and *in vivo* studies suggest that SIM may also dimerize with an ARNT-like protein to regulate enhancer sequences present in the *sim*, *slit* and *Toll* structural genes. The *Drosophila* PER protein plays a role in the maintenance of circadian rhythms. PER has been shown to form heterotypic interactions with a second *Drosophila* protein, TIM, *in vivo*, and homotypic interactions with the ARNT molecule *in vitro*.

The distinguishing characteristic of these proteins is a 200-300 stretch of amino acid sequence similarity known as a PAS (PER/ARNT/SIM) domain. In the AHR, the PAS domain has been shown to encode sites for agonist binding, surfaces to support heterodimerization with other PAS domains, as well as surfaces that form tight interactions with Hsp90. In addition to the PAS domain, the AHR, ARNT and SIM also harbor a bHLH (basic helix-loop-helix) motif that plays a primary role in dimer formation. The bHLH motif is found in a variety of transcription factors that utilize homotypic interactions to regulate various aspects of cell growth and differentiation. Dimerization specificity is dictated by sequences within both the bHLH and determinants within secondary interaction surfaces, such as the "leucine zipper or PAS domains. Interestingly, these dimerization

surfaces also appear to restrict pairing to within a given bHLH protein superfamily, thus minimizing crosstalk between important cellular pathways.

Because other bHLH protein families utilize multiple homotypic interactions to provide fine control in the regulation of certain gene batteries, it is possible that additional bHLH-PAS proteins exist in the mammalian genome and that a subset of these proteins might dimerize with either the AHR or ARNT. However, prior to the present invention, the AHR and ARNT were the only mammalian bHLH-PAS proteins that had been identified. Accordingly, a need exists to identify and characterize other bHLH-PAS domain proteins, particularly those that are novel receptors for drugs, or are AHR or ARNT binding partners. Such molecules would find broad utility as research tools in elucidating environmentally and developmentally controlled signal transduction pathways, and also as diagnostic and therapeutic agents for detection and control of such pathways.

SUMMARY OF THE INVENTION

This invention provides isolated nucleic acids and proteins which are new and distinct members of the bHLH-PAS superfamily of transcription regulators. These "MOPs" (members of PAS) are useful for a wide variety of research, diagnostic and therapeutic applications, as described in greater detail herein.

According to one aspect of the invention, isolated nucleic acid molecules are provided that include an open reading frame encoding a protein selected from the group consisting of: MOP2, MOP3, MOP4, MOP5, MOP6 MOP7, MOP8 and MOP9. In preferred embodiments, the open reading frame encodes a protein having an amino acid sequence substantially the same as a sequence selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 SEQ ID NO:17 and SEQ ID NO:18. The nucleic acid

molecules of the invention preferably comprise sequences substantially the same as a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 SEQ ID NO:8 and SEQ ID NO:9.

According to another aspect of the invention, isolated MOP proteins are provided, which are products of expression of part or all of the open reading frames of the aforementioned nucleic acid molecules.

According to another aspect of the invention, recombinant DNA molecules are provided, which comprise MOP encoding nucleic acid molecules, operably linked to vectors for transforming cells. Cells transformed with those recombinant DNA molecules are also provided, as well as cellular assay systems utilizing those recombinant molecules.

According to another aspect of the invention, oligonucleotides between about 10 and about 100 nucleotides in length are provided, which specifically hybridize with portions of the MOP-encoding nucleic acid molecules.

According to another aspect of the invention, antibodies are provided which are immunologically specific for part or all of any of the MOP2-MOP8 proteins of the invention.

According to another aspect of the invention, assays and other methods of using the aforementioned MOP nucleic acids, proteins and immunospecific antibodies are provided.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1. Schematic representation of a generic bHLH-PAS member and the corresponding region where EST "hits" occurred. *Top,* schematic of a generic bHLH-PAS family member. The *hatched box* represents the bHLH region, the *overlined* area represents the PAS domain with the characteristic "A" and "B" repeats in *white*.

The variable C terminus is *boxed in white*. A *bold line* representing the region in a generic bHLH-PAS member where the homology occurs is indicated next to the original Gen Bank™ accession number for each identified EST (MOP1=T10821, MOP2 = T70415, MOP3 = T77200 and F06906, MOP4 = R58054, MOP5 = R67292; see Table 1).

FIGURE 2. Amino acid sequence and multiple alignment of the PAS domains. The amino acid sequence including a CLUSTAL alignment of the bHLH-PAS domains is depicted. The CLUSTAL alignment was performed using the MEGALIGN program (DNASTAR, Madison, WI) with a PAM250 weight table using the following parameters: Ktuple = 1, Gap Penalty = 3, Window = 5. Amino acid boundaries for the residues encompassing the bHLH and PAS domains of the MOPs were defined based on previous observations. The bHLH domain is *boxed,* while the basic region is specified by a *vertical line*. The PAS domain is *underlined,* while the "A" and "B" repeats of the PAS domain are *boxed*. Consensus (60%) residues in the PAS domain are denoted with an *asterisk*.

FIGURE 3. Yeast two-hybrid analysis. *In vivo* interaction of MOPs with dioxin signaling pathway. Fig. 3A, schematic of AHR, ARNT, and LexA fusion constructs. Panel shows a schematic of the AHR, with the PAS domain (*black*) with the characteristic *"A"* and *"B"* repeats (*white*), the bHLH domain (*striped*), and the variable C terminus (*white*). The transcriptionally active glutamine rich domain is indicated with *"Q"* (*shaded box*). LexA fusion proteins are indicated with the N terminus of LexA DNA-binding protein fused to bHLH-PAS domains of MOPs 1-4, and ARNT. The LexAAHR construct contains the bHLH-PAS domains and the C terminus minus the transcriptionally active Q-rich region (see Example 1, "Materials and Methods"). Fig. 3B, relative interaction of LexA fusion proteins with the AHR or ARNT. Galacto-Light assays were performed on yeast extracts prepared from colonies expressing LexAMOPs, LexAAHR, or LexAARNT and ARNT or AHR

in the presence and absence of 1 μM βNF. Assays were performed in triplicate, and then the relative light units normalized to the LexAAHR-ARNT + βNF condition internally (set as 100%). The *stippled bars* represent LexA fusion proteins co-expressed with the full-length AHR in the presence of 1 μM βNF, the *striped bars* represent the fusion proteins co-expressed with the full length AHR in the absence of ligand, the *shaded bars* indicate the fusion proteins co-expressed with the full-length ARNT, and the *open bar* indicates LexAAHR co-expressed with full-length ARNT in the presence of 1 μM βNF.

FIGURE 4. Schematic comparison of homology of PAS family members. A dendrogram was prepared from the primary amino acid CLUSTAL alignment above using the MEGALIGN program. The CLUSTAL alignment was performed using the MEGALIGN program (DNASTAR, Madison, WI) with PAM250 weight table using the following parameters: Ktuple =1, Gap Penalty = 3, Window = 5. Amino acid boundaries for the residues encompassing the bHLH and PAS domains of the MOPs were defined based on previous observations. The amino acid boundaries are as follows: huMOP1/HIF1 (91-342), huMOP2 (90-342), huMOP3 (148-439), huMOP4 (87-350), huMOP5 (32-296), huAHR (117-385), huARNT (167-464), drSIM (82-356), drPer (232-496) bsKINA (27-248), huSRC-1(115-365), muARNT2 (141-437, muSIM1 (83-331), muSIM2 (83-332), drSIMILAR (174-419), and drTRH (145-471). The scale at the bottom indicates number of amino acid residue substitutions. PAS family members that interact with HSP90, interact with the AHR, and interact with ARNT by the coimmunoprecipitation method are indicated by a +, whereas members that do not interact are indicated by a -. An α denotes a bHLH-PAS member whose cDNA is not complete. Note that these interactions occur *in vitro* and may or may not be physiologically relevant. Where appropriate, the reference is included in parentheses.

FIGURE 5. Interaction panel of LexAbHLH-PAS fusion proteins with full-length MOP3 and ARNT. Fig. 5A: Schematic representation of the LexAbHLHPAS "bait" and the full-length "fish." The bHLH and PAS domains are boxed. The "A" and "B" repeats of the PAS domains are indicated. The transactivation domain of the full-length "fish" is indicated. Fig. 5B: LexA fusion protein plasmids containing the bHLH-PAS domains of HIF1α, HIF2α, MOP3, MOP4, AHR, ARNT, and CLOCK were coexpressed with plasmids harboring full-length MOP3 and ARNT (see Materials and Methods). LexAAHR interactions were assayed on plates containing 1 μM β-naphthoflavone. After incubation, an 5-bromo-4-chloro-3-indolyl 13-β-galactoside overlay assay was performed. ++, A strong interaction, turning blue within 2 hr; +, a weaker interaction, turning blue between 8 hr and overnight; and -, a negative interaction after overnight incubation. The experiment was performed three times with identical results.

FIGURE 6. The consensus DNA binding site for MOP3-MOP4 heterodimer *in vitro*. Ten selected DNA sequences bound by the MOP3-MOP4 complex are indicated with the E-box core boxed. Underneath, the M34 consensus is indicated. Nucleotide positions relative to the E-box core are shown. Bases in uppercase are randomer derived, while bases in lower case are primer derived.

FIGURE 7. Cloning of MOP7. The positions of the original EST clone (AA028416) and RACE products are shown as dark lines with the mMOP7 ORF shown as an open box. The PCR primers used are posted below the corresponding fragments and the plasmid numbers are marked on the side. The GenBank Accession Number for mouse MOP7 cDNA is AF060194.

FIGURE 8. The splicing site within mouse MOP7 ORF are compared with those previously reported for mHIF1α and hHIF2α. The numbers of amino acids at which the splicing occurs are marked underneath the sequence.

The conserved splicing sites are defined as the splicing sites of HIF1α and HIF2α that are within one amino acid of the corresponding MOP7 splicing site on the aligned sequence map using CLUSTAL method. These sites are marked with lines between different ORFs (see GenBank Accession Numbers AF079140-079153 for detailed sequences of mMOP7 splice sites.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Various terms relating to the biological molecules of the present invention are used hereinabove and also throughout the specifications and claims. The terms "substantially the same," "percent similarity" and "percent identity" are defined in detail below.

With reference to nucleic acids of the invention, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

With respect to RNA molecules of the invention, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to protein, the term "isolated protein" or "isolated and purified protein" is sometimes

used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

With respect to antibodies of the invention, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to single stranded nucleic acids, particularly oligonucleotides, the term "specifically hybridizing" refers to the association between two single-stranded nucleotide molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

The term "promoter region" refers to the transcriptional regulatory regions of a gene. In the

present invention, the use of SV40, TK, Albumin, SP6, T7 gene promoters, among others, is contemplated.

The term "selectable marker gene" refers to a gene encoding a product that, when expressed, confers a selectable phenotype such as antibiotic resistance on a transformed cell.

The term "reporter gene" refers to a gene that encodes a product which is easily detectable by standard methods, either directly or indirectly.

The term "operably linked" means that the regulatory sequences necessary for expression of the coding sequence are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement of transcription units and other transcription control elements (e.g. enhancers) in an expression vector.

II. Characterization of MOPS 1-9

Our hypothesis in accordance with the present invention was that additional bHLH-PAS proteins are encoded in the mammalian genome and that some of these proteins are involved in mediating the pleiotropic response to potent AHR agonists like TCDD. It has been observed that other bHLH superfamilies employ multiple dimeric partnerships to control complex biological processes, such as myogenesis (MyoD/myogenin), cellular proliferation (Myc, Max, Mad) and neurogenesis (*achaete-scute/daughterless*). The observation that bHLH proteins often restrict their dimerization to within members of the same gene family (i.e., "homotypic interactions") and that this restriction may occur as the result of constraints imposed by both primary (e.g., bHLH) and secondary dimerization surfaces (e.g., leucine zippers and PAS), prompted us to screen for additional bHLH-PAS proteins and test each protein for its capacity to interact with either the AHR or ARNT. The ultimate

objective was to identify MOPs that were physiologically relevant partners of either the AHR or ARNT *in vivo*. Our prediction was that such proteins might respond to or modulate the AHR signaling pathway or other signaling pathways involving ARNT.

To rapidly identify expressed genes, the "expressed sequence tag" (EST) approach was developed, whereby a cDNA library is constructed and randomly selected clones are sequenced from both vector arms (Adams et al., Science 252: 1651-1656, 1991). These partial sequences, generally 200-400 bp, are deposited in a number of computer databases that can be readily analyzed using a variety of search algorithms. As of 1996, the I.M.A.G.E. Consortium has deposited over 300,000 human ESTs, generated from different tissues and developmental time periods into publicly accessible databases, identifying approximately 40,000 unique cDNA clones (Lennon et al., Genomics 33: 151-152, 1996). The availability of these sequences and plasmids harboring their corresponding cDNA clones provided a means by which to identify novel members of the bHLH-PAS family by nucleotide homology screening of available EST databases.

At the time this invention was initiated, the human AHR and ARNT and the drosophila SIM and PER were the only PAS protein that had been described. Therefore, we used the nucleotide sequences encoding their PAS domains as query sequences in BLASTN searches of the available EST databases. Using this strategy in an iterative fashion and confirming each hit with a reverse BLASTX search, we have identified eight cDNAs referred to herein as members of the PAS superfamily, or "MOPs". Using PCR, we were able to obtain the complete ORFs of MOPs 1-4, and extensive but incomplete ORFs of MOP5. We have also identified four more MOPs, MOPs 6, 7, 8 and 9, and obtained their complete ORFs.

While MOPs 1-5 were being characterized, Wang and colleagues identified two factors involved in

cellular response to hypoxia, HIF1α and HIF1β. These proteins are identical to MOP1 and ARNT, respectively (Wang et al., Proc. Natl. Acad. Sci. USA 92: 5510-5514, 1995). Thus, of the nine MOPs we have cloned, seven have not been previously characterized. For consistency herein, we describe MOP1 extensively, and describe heretofore undisclosed methods of using MOP1.

The experimental approach taken in accordance with the present invention has significantly expanded the number of known members of the emerging bHLH-PAS superfamily of transcriptional regulators. Along with the MOPs described herein, five additional mammalian bHLH-PAS proteins have been identified, HIF1α (MOP1, as described above), SIM1, SIM2, ARNT2, and SRC-1 (Wang et al., 1995, *supra*; Hirose et al., Mol. Cell. Biol. 16: 1706-1713, 1996; Fan et al., Mol. Cell. Neurosci. 7: 1-16, 1996; Ema et al., Mol. Cell. Biol. 16: 5865-5875, 1996; Chen et al., Nat. Genet. 10: 9-10, 1995; and Kamei et al., Cell 85: 403-414, 1996). To compare amino acid sequences of these proteins, we performed a CLUSTAL alignment with the bHLH-PAS domains of MOPs 1-5 and all the known family members using a PAM250 residue weight table (Higgins & Sharp, Gene (Amst.) 73: 237-244, 1988). The two most related members were MOP1/HIF1α and MOP2, which shared 66% identity in the PAS domain. A comparison of these two proteins reveals only a single amino acid difference in the basic region and 83% identity in the HLH region. This sequence similarity is in agreement with our contention (discussed in Example 1) that MOP1/HIF1α and MOP2 function analogously, interacting with the same heterodimeric partners and binding similar enhancer sequences *in vivo*. A comparison of MOP3 and ARNT and a comparison of MOP5 and SIM reveal 40% and 38% identity in the PAS domain, respectively. The basic regions of MOP3 and ARNT have only three substitutions, while the HLH domains share 66% identity, again suggesting that the two proteins may regulate

similar or identical enhancer sequences (half sites).

A CLUSTAL alignment of the C-termini of MOPs 1-5 and the previously identified PAS members demonstrated that these regions are not well conserved (data not shown) (Burbach et al., Proc. Natl. Acad. Sci. USA 89: 8185-8189, 1992). This lack of conservation may indicate that the C-termini of these genes have divergent functions, or that the functions harbored in the C-termini can be accomplished by a variety of different sequences. For example, the C-termini of the AHR, ARNT, and SIM all harbor potent transactivation domains, yet display little sequence homology.

To characterize the evolutionary and functional relationships of these proteins, we performed a parsimony analysis to identify functionally related subsets. A dendrogram representing the primary amino acid relationship between the PAS domains of these proteins is illustrated in Figure 4. This schematic suggests that major groups exist for eukaryotic PAS family members. The AHR, drSIMILAR, MOP1/HIF1α, MOP2, drTRACHEALESS, MOP5, and SIM exist in one group, ARNT, muARNT2, MOP3, and MOP4 in another and PER and huSRC-1 exist in their own groups. Interestingly, this pattern reflects what is known functionally about the existing PAS members. The AHR, SIM, MOP1/HIF1α and MOP2 have all been shown to heterodimerize with the ARNT molecule and bind DNA. Additionally, the AHR and SIM are known to interact with HSP90, a chaperonin protein necessary for the signaling of the AHR and a number of steroid receptor family members in response to ligand. Based on these groupings, MOP5 may also be an ARNT-interacting protein and a candidate for interacting with Hsp90 and being activated by small molecule ligands. The observation that ARNT has been shown to be capable of forming DNA binding homodimers and as heterodimers with a number of previously identified members of the bHLH-PAS family (at least *in vitro*), suggests that it plays a role in a

number of biological processes. Based on their similarity with ARNT, MOP3 and MOP4 may be candidates for binding DNA as homodimers, or for interacting with multiple bHLH-PAS members, possibly from the AHR group.

In addition to the relevance of the above data to TCDD signaling, they also reveal additional factors important to cellular responses to hypoxic stress. HIF1α/MOP1 and MOP2 appear to share a common dimeric partner - ARNT, and are capable of regulating a common battery of genes. This notion is supported by three lines of evidence: (1) both MOP1 and MOP2 interact with ARNT as defined by coimmunoprecipitation or two-hybrid assay; (2) they have similar DNA half-site specificities when complexed with ARNT; and (3) they are both transcriptionally active from TACGTG enhancers *in vivo*. The observation that HIF1α/MOP1 and MOP2 have markedly different tissue distributions suggests that these two proteins may be regulating similar batteries of genes in response to different environmental stimuli. Alternatively, these proteins may be involved in restricting expression of certain groups of genes regulated by TACGTG-dependent enhancers. Finally, it is now known that MOP2 and MOP7 are subunits of a "HIF1-like" complex (i.e. a "HIF2α" and a "HIF3α, respectively) that regulates hypoxia responsive genes in distinct sets of tissues.

From the foregoing discussion, it can be seen that, while the MOPs share certain common features among themselves and with other new members of the bHLH-PAS superfamily, each of MOPs 2 - 9 is a distinctive and unique member of that family. cDNA and deduced amino acid sequences for each of MOPs 1 - 9 is set forth at the end of this specification. General features of each MOP are summarized below. In addition, MOPs 1-5 are described in great detail in Example 1, MOP3 is specifically described in Example 2 and MOP 7 is described in Example 3.

MOP1: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP1 are set forth herein as SEQ ID NOS: 1 and 10, respectively. The cDNA includes a complete coding sequence for MOP1. As discussed above, MOP1 is known more commonly in the literature as HIF (Hypoxia-Inducible Factor)-1α (Wang et al., 1995, *supra*). The factor is induced by low oxygen. It interacts with HSP90 and with ARNT (AHR's binding partner). The ARNT-dimerized factor regulates expression of erythropoietin, among other genes.

MOP2: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP2 are set forth herein as SEQ ID NOS: 2 and 11, respectively. The cDNA includes a complete coding sequence for MOP2. MOP2 appears to be related structurally and functionally to MOP1. Similar to MOP1, MOP2 interacts with ARNT, but not AHR, and drives transcription in its ARNT-dimerized form. Unlike MOP1, MOP2 does not appear to interact significantly with HSP90. MOP2 is induced by low oxygen and may be involved in hypoxia responses in different cells and tissues than is MOP1. MOP2 is sometimes referred to herein as HIF2α.

MOP3: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP3 are set forth herein as SEQ ID NOS: 3 and 12, respectively. The cDNA includes a complete coding sequence for MOP3. MOP 3 and MOP 4 are related to each other as binding partners, analogous to ARNT and AHR, respectively. As described in greater detail in Example 2, in addition to being a specific partner for MOP4, MOP3 is a general dimerization partner for a subset of the bHLH/PAS superfamily of transcriptional regulators. MOP3 interacts with MOP4, CLOCK, HIF1α and HIF2α. The MOP3-MOP4 heterodimer binds a CACGTGA-containing DNA element. Moreover, MOP3-MOP4 and MOP3-CLOCK complexes bind this element in COS-1 cells and drive transcription from a linked luciferase reporter gene. A high-affinity DNA binding site has also been deduced for a MOP3-HIF1α complex (TACGTGA). MOP3-HIF1α

and MOP3-HIF2α heterodimers bind this element, drive transcription, and respond to cellular hypoxia.

MOP3 also binds HSP90, and may be conditionally activated (like AHR) depending on whether it is bound to HSP90 (see Example 1) (of the MOP3/MOP4 dimerization pair, one appears to be conditionally activated, but as yet it is unclear which one). Evidence from *Drosophila* and rat suggest that MOP3 (cycle/ bMAL1b) is regulated in a circadian manner.

MOP3 expression appears to be controlled by alternate 5' promoter regions. MOP3 mRNA expression overlaps in a number of tissues with each of its four potential partner molecules *in vivo*.

MOP4: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP4 are set forth herein as SEQ ID NOS: 4 and 13, respectively. The cDNA includes an apparently complete coding sequence for MOP4. MOP4 appears to be a human ortholog of a recently identified murine gene called *"Clock"*, for its involvement in circadian rhythms (King et al., Cell 89: 641-653). MOP4 also interacts with HSP90 and, as discussed above, is the dimerization partner of MOP3, and may be conditionally activated. MOP4 appears to be localized in the cytoplasm.

MOP5: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP5 are set forth herein as SEQ ID NOS: 5 and 14, respectively. The cDNA includes a partial coding sequence for MOP5; however a complete coding sequence for MOP5 has become publicly available subsequent to the making of the present invention (GenBank Accession No. U77968, submitted 11/11/96, published 1/21/97 by Zhou et al., Proc. Natl. Acad. Sci. USA 94: 713-718).

MOP6: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP6 (of human origin) are set forth herein as SEQ ID NOS: 6 and 15, respectively. The cDNA includes a complete coding sequence for MOP6.

The nucleotide sequence of MOP6 is fairly unique. It is most similar in the 5' region to the bHLH-PAS member *trachealess*, which suggests that MOP6 may be a regulator (developmental or otherwise) of hypoxia. Functional data shows that MOP6 forms a partnership with ARNT and drives a hypoxia responsive element.

MOP7: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP7 are set forth herein as SEQ ID NOS: 7 and 16, respectively. The cDNA includes a complete coding sequence for MOP7. In accordance with this invention, MOP7 has been characterized as a new hypoxia-inducible factor, and therefore is sometimes referred to herein as HIF3α. The expression profile of MOP7 is as follows: testis, thymus > [lung, brain, heart, liver, skeletal muscle] > [skin, stomach, small intestine, kidney]. This expression profile is distinct from any of MOP1, MOP2, MOP3, AHR and ARNT, suggesting a different functional role for MOP7. MOP7 is most closely related to MOP1/HIF1α and MOP2 (HIF2α), as described in greater detail in Example 3. Accordingly, MOP7 is likely to regulate the same genes as does HIF1α and HIF2α, as evidenced by its dimerization with the same partners (ARNT, MOP3) and recognition of the same core response element. This, combined with the unique tissue-specific expression of MOP7 suggests that it may have a functional role associated with response to low oxygen in the tissues in which it is expressed.

MOP8: The nucleotide and deduced amino acid sequences of a cDNA encoding MOP8 are set forth herein as SEQ ID NOS: 8 and 17, respectively. The cDNA includes a complete coding sequence for MOP8. Like MOP4 and MOP3, MOP8 may be involved in regulation of circadian rhythm. MOP8 shows sequence similarity to other genes involved in the circadian pathway (human PER, *Drosophila* PER, human RIGUI).

MOP9: The nucleotide and deduced amino acid sequence of a cDNA encoding MOP9 are set forth herein as SEQ ID NOS: 9 and 18, respectively. Two ESTs (GenBank AA577389, AA576971) corresponding to a novel bHLH-PAS protein homologous to MOP3/bMAL1 were identified by TBLASTN searches of the Drosophila homolog of MOP3. Upon characterization, these clones were revealed to be truncated, and one of which appeared to be a splice variant. The cDNA was cloned from human brain mRNA, and alternative 5' splicing was found probably reflecting multiple promoters. A BLASTX search of the MOP 9 sequence reveals that it displays extended homology to MOP3 (E-154). These data suggest that MOP9 also pairs with CLOCK and MOP4 and binds an E-box element with flanking region specificity.

Although specific MOP clones are described and exemplified herein, this invention is intended to encompass nucleic acid sequences and proteins from humans and other species that are sufficiently similar to be used interchangeably with the exemplified MOP nucleic acids and proteins for the purposes described below. It will be appreciated by those skilled in the art that MOP-encoding nucleic acids from diverse species, and particularly mammalian species, should possess a sufficient degree of homology with human MOPs so as to be interchangeably useful in various applications. The present invention, therefore, is drawn to MOP-encoding nucleic acids and encoded proteins from any species in which they are found, preferably to MOPs of mammalian origin, and most preferably to MOPs of human origin. Additionally, in the same manner that structural homologs of human MOPs are considered to be within the scope of this invention, functional homologs are also considered to be within the scope of this invention.

Allelic variants and natural mutants of SEQ ID NOS: 1-9 or 10-17 are likely to exist within the human

genome and within the genomes of other species. Because such variants are expected to possess certain differences in nucleotide and amino acid sequence, this invention provides isolated MOP-encoding nucleic acid molecules having at least about 65% (and preferably over 75%) sequence homology in the coding region with the nucleotide sequences set forth as SEQ ID NOS: 1-9 (and, most preferably, specifically comprising the coding regions of any of SEQ ID NOS: 1-9). This invention also provides isolated MOPs having at least about 75% (preferably 85% or greater) sequence homology with the amino acid sequence of SEQ ID NOS: 10-18. Because of the natural sequence variation likely to exist among the MOPs and nucleic acids encoding them, one skilled in the art would expect to find up to about 25-35% nucleotide sequence variation, while still maintaining the unique properties of the MOPs of the present invention. Such an expectation is due in part to the degeneracy of the genetic code, as well as to the known evolutionary success of conservative amino acid sequence variations, which do not appreciably alter the nature of the protein. Accordingly, such variants are considered substantially the same as one another and are included within the scope of the present invention.

For purposes of this invention, the term "substantially the same" refers to nucleic acid or amino acid sequences having sequence variation that do not materially affect the nature of the protein. With particular reference to nucleic acid sequences, the term "substantially the same" is intended to refer to the coding region and to conserved sequences governing expression, and refers primarily to degenerate codons encoding the same amino acid, or alternate codons encoding conservative substitute amino acids in the encoded polypeptide. With reference to amino acid sequences, the term "substantially the same" refers generally to conservative substitutions and/or variations

in regions of the polypeptide not involved in determination of structure or function of the protein. The terms "percent identity" and "percent similarity" are also used herein in comparisons among amino acid sequences. These terms are intended to be defined as they are in the UWGCG sequence analysis program (Devereaux et al., Nucl. Acids Res. 12: 387-397, 1984), available from the University of Wisconsin.

The following description sets forth the general procedures involved in practicing the present invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures, such as those set forth in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989) (hereinafter "Sambrook et al.") or Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998) (hereinafter "Ausubel et al.") are used.

III. Preparation of MOP Nucleic Acid Molecules, MOP Proteins and Anti-MOP Antibodies

A. Nucleic Acid Molecules

Nucleic acid molecules encoding the MOPs of the invention may be prepared by two general methods: (1) They may be synthesized from appropriate nucleotide triphosphates, or (2) they may be isolated from biological sources. Both methods utilize protocols well known in the art.

The availability of nucleotide sequence information, such as a full length cDNA having any of SEQ ID NOS: 1-9, enables preparation of an isolated nucleic acid molecule of the invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramadite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct may be purified according to methods known in the art, such as high performance liquid

chromatography (HPLC). Long, double-stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current oligonucleotide synthetic methods. Thus, for example, a several-kilobase double-stranded molecule may be synthesized as several smaller segments of appropriate complementarity. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the presence of DNA ligase to construct an entire double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an appropriate vector.

Nucleic acid sequences encoding MOPs may be isolated from appropriate biological sources using methods known in the art. In a preferred embodiment, cDNA clones are isolated from libraries of human origin. In an alternative embodiment, genomic clones encoding MOPs may be isolated. Alternatively, cDNA or genomic clones encoding MOPs from other species, preferably mammalian species, may be obtained.

In accordance with the present invention, nucleic acids having the appropriate level sequence homology with the coding regions of any of Sequence I.D. Nos. 1-9 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 μg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS; (4) 2

hours at 42-65°in 1X SSC and 1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology (Sambrook et al., 1989):

$T_m = 81.5°C + 16.6\text{Log}[Na+] + 0.41(\%\ G+C) - 0.63\ (\%\ \text{formamide}) - 600/\#\text{bp in duplex}$

As an illustration of the above formula, using [N+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a preferred embodiment, clones are maintained in plasmid cloning/expression vector, such as pGEM-T (Promega Biotech, Madison, WI) or pBluescript (Stratagene, La Jolla, CA), either of which is propagated in a suitable *E. coli* host cell.

MOP nucleic acid molecules of the invention include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides (sense or antisense strands of DNA or RNA) having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of the cDNA having any of SEQ ID NOS: 1-9. Such oligonucleotides are useful as probes for detecting MOP genes or mRNA in test samples of cells, tissue or other biological sources, e.g. by PCR amplification, or for the positive or negative regulation of expression of MOP genes at or before translation of the mRNA into proteins.

B. Proteins

MOP proteins of the present invention may be prepared in a variety of ways, according to known methods. The proteins may be purified from appropriate sources, e.g., cultured or intact cells or tissues.

Alternatively, the availability of nucleic acids molecules encoding MOPs enables production of the MOP proteins using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such a pSP64 or pSP65 for *in vitro* transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocytes. *In vitro* transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison, Wisconsin or BRL, Rockville, Maryland.

According to a preferred embodiment, larger quantities of MOP proteins may be produced by expression in a suitable procaryotic or eucaryotic system. For example, part or all of a DNA molecule, such as any of the cDNAs having SEQ ID NOS: 1-9, may be inserted into a plasmid vector adapted for expression in a bacterial cell (such as *E. coli*) or a yeast cell (such as *Saccharomyces cerevisiae*), or into a baculovirus vector for expression in an insect cell. Such vectors comprise the regulatory elements necessary for expression of the DNA in the host cell, positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The MOPs produced by gene expression in a recombinant procaryotic or eucyarotic system may be purified according to methods known in the art. In a preferred embodiment, a commercially available expression/secretion system can be used, whereby the recombinant protein is expressed and thereafter secreted

from the host cell, to be easily purified from the surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein. Such methods are commonly used by skilled practitioners. The MOP proteins of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures.

The present invention also provides antibodies capable of immunospecifically binding to MOP proteins of the invention. Polyclonal or monoclonal antibodies directed toward any of MOPs 1-9 may be prepared according to standard methods. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. In a preferred embodiment, antibodies have been prepared, which react immunospecifically with various epitopes of the MOPs.

Polyclonal or monoclonal antibodies that immunospecifically interact with MOPs can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules. Other uses of anti-MOP antibodies are described below.

IV. Uses of MOP-Encoding Nucleic Acids, MOP Proteins and Anti-MOP Antibodies

A. MOP-Encoding Nucleic Acids

MOP-encoding nucleic acids may be used for a variety of purposes in accordance with the present invention. MOP-encoding DNA, RNA, or fragments thereof may be used as probes to detect the presence of and/or expression of genes encoding MOPs. Methods in which MOP-encoding nucleic acids may be utilized as probes for such assays include, but are not limited to: (1) *in situ*

hybridization; (2) Southern hybridization (3) northern hybridization; and (4) assorted amplification reactions such as polymerase chain reactions (PCR). In addition, recombinant cellular assay systems to examine signal transduction pathways in which the MOPs are involved are described below.

The MOP-encoding nucleic acids of the invention may also be utilized as probes to identify related genes either from humans or from other species. As is well known in the art, hybridization stringencies may be adjusted to allow hybridization of nucleic acid probes with complementary sequences of varying degrees of homology. Thus, MOP-encoding nucleic acids may be used to advantage to identify and characterize other genes of varying degrees of relation to the respective MOPs, thereby enabling further characterization the AHR or related signaling cascades. Additionally, they may be used to identify genes encoding proteins that interact with MOPs (e.g., by the "interaction trap" technique, or modifications thereof, as described in Example 1), which should further accelerate elucidation of these cellular signaling mechanisms.

Nucleic acid molecules, or fragments thereof, encoding MOPs may also be utilized to control the production of the various MOPs, thereby regulating the amount of protein available to participate in cellular signaling pathways. In one embodiment, the nucleic acid molecules of the invention may be used to decrease expression of certain MOPs in cells. In this embodiment, full-length antisense molecules are employed which are targeted to respective MOP genes or RNAs, or antisense oligonucleotides, targeted to specific regions of MOP-encoding genes that are critical for gene expression, are used. The use of antisense molecules to decrease expression levels of a pre-determined gene is known in the art. In a preferred embodiment, antisense oligonucleotides are modified in various ways to increase

their stability and membrane permeability, so as to maximize their effective delivery to target cells *in vitro* and *in vivo*. Such modifications include the preparation of phosphorothioate or methylphosphonate derivatives, among many others, according to procedures known in the art.

In another embodiment, the transcription regulation activity of bHLH-PAS homodimers or heterodimers involving MOPs may be blocked by genetically engineering a target cell to express a defective MOP - specifically one that has been modified to be unable to bind DNA. When the defective MOP dimerizes, the dimer is also unable to bind DNA, and therefore is unable to carry out its transcriptional regulatory function.

In another embodiment, overexpression of various MOPs is induced, which can lead to overproduction of a selected MOP. Overproduction of MOPs may facilitate the isolation and characterization of other components involved in protein-protein complex formation occurring during the MOP-related signal transduction in cells.

As described above, MOP-encoding nucleic acids are also used to advantage to produce large quantities of substantially pure MOP proteins, or selected portions thereof.

B. MOP Proteins and Anti-MOP Antibodies

Purified MOPs, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies which also may serve as sensitive detection reagents for the presence and accumulation of MOPs (or complexes containing the MOPs) in cultured cells or tissues or in intact organisms. Recombinant techniques enable expression of fusion proteins containing part or all of a selected MOP protein. The full length protein or fragments of the protein may be used to advantage to generate an array of monoclonal or polyclonal antibodies specific for various epitopes of the protein, thereby providing even greater sensitivity for

detection of the protein in cells or tissue.

Polyclonal or monoclonal antibodies immunologically specific for a MOP may be used in a variety of assays designed to detect and quantitate the protein. Such assays include, but are not limited to: (1) flow cytometric analysis; (2) immunochemical localization of a MOP in cells or tissues; and (3) immunoblot analysis (e.g., dot blot, Western blot) of extracts from various cells and tissues. Additionally, as described above, anti-MOPs can be used for purification of MOPs (e.g., affinity column purification, immunoprecipitation).

C. Recombinant Cells and Assay Systems

Genetically engineered cells, such as yeast cells or mammalian cells, may be produced to express any one, or a combination, of MOPs described herein. Such cells can be used to evaluate the binding interactions between MOPs, or between a MOP and another member of the bHLH-PAS superfamily (e.g., AHR, ARNT), and the requirement for homodimerization or heterodimerization of the MOPs for initiation of transcriptional control of a reporter gene driven by appropriate enhancer elements. In addition, such recombinant cells can be used to study the effect of external stimuli, such as hypoxia or TCDD, on activation of a selected MOP, or they can be used to screen panels of drugs for control of MOP-involved signal transduction pathways. U.S. Patent No. 5,650,283 to Bradfield et al., the disclosure of which is incorporated herein by reference, describes recombinant cellular systems and assays for detecting agonists to the AHR. These materials and methods may be used similarly to design recombinant systems for evaluating any of MOP1-MOP8, in the presence or absence of an external stimulant.

Appropriate yeast cells for production of such recombinant systems include *Saccharomyces cerevisiae* and *Saccharomyces pombe*. Yeast strains carrying endogenous functional HSPs may be utilized (e.g., A303 obtained from

Rick Gaber, Northwestern University, or commercially available equivalents). Yeast strains in which the genes encoding HSPs have been disrupted may also be utilized (e.g., GRS4, obtained from Susan Lindquist, University of Chicago), affording an opportunity to examine the relationship of various MOPs to HSPs.

Appropriate mammalian cells for production of such recombinant systems include COS, Hep3b, HepGr and Hepalclc7 cells, among others.

In one type of assay where the MOP signal transduction pathway is affected by an external stimulus (i.e. an agonist such as TCDD in the AHR-ARNT system, or cobalt chloride in the MOP1/HIF1α-ARNT system), an appropriate cell can be transformed with an expression plasmid expressing a full length agonist receptor MOP, along with its dimerization partner (if the MOP forms heterodimers) and a reporter plasmid expressing a reporter gene, such as *LacZ* or luciferin, which is driven by an appropriate enhancer element. The presence or potency of a selected agonist may be determined by its ability to activate transcription of the reporter gene in the recombinant system.

In another embodiment, a recombinant system that does not rely on heterodimerization can be constructed. In this case, a cell is transformed with an expression plasmid expressing a chimeric agonist-sensitive MOP, along with a reporter plasmid expressing a reporter gene driven by a suitable promoter. The chimeric MOP is modified to replace the heterodimerization domains (i.e. the bHLH-PAS domain) with a DNA binding domain, such as LexA or Gal4. Such chimeras will homodimerize and activate transcription of genes positioned downstream of LexA or Gal4 binding sites engineered into the reporter plasmid.

In a preferred embodiment, described in detail in Example 1, a modified yeast "two hybrid" system is used to assess binding interactions between MOPs (and other bHLH-PAS proteins) and the subsequent initiation of

transcriptional control. For instance, as described in Example 1, fusion proteins were constructed in which the DNA binding domain of the bacterial repressor, LexA, was fused to the bHLH-PAS domains of the MOP proteins. Interactions were tested by cotransformation of each LexAMOP construct with either the full length AHR or ARNT into the L40 yeast strain, which harbors an integrated *lacZ* reporter gene driven by multiple LexA operator sites. In this system, LexAMOP fusions which interact with AHR or ARNT drive expression of the *lacZ* reporter gene. The effect of various agonists on reporter gene expression can also be evaluated using this system.

Any one or more of the aforementioned recombinant cell systems and assays can be used to screen panels of drugs for their effect on specific signal transduction pathways. For instance, recombinant systems employing any or MOPs 1, 2, 6 or 7 may be used to screen for drugs that stimulate red blood cell synthesis, angiogenesis or glucose metabolism.

Recombinant systems employing any of MOPs 3, 4, 8 or 9 may be used to screen for drugs that modify circadian rhythms. In connection with this embodiment, as described in greater detail in Example 2, we have determined the binding sequence for the MOP3/MOP4 heterodimer, and have constructed the following recombinant plasmids: PL833, a MOP3 expression vector for mammalian cells; PL834, a MOP4 expression vector for mammalian cells; and PL880, a reporter plasmid (expressing luciferase) driven by the MOP3/MOP4 consensus enhancer sequence GCA_CACGTG_ACC. When the three plasmids are introduced into a mammalian cell, the reporter gene responds to the presence of the MOP3/MOP4 dimer. This system is used in a high throughput microwell assay to screen for compounds that are specific activators or inhibitors of these transcription factors. A similar system has been established for MOP7 (HIF3α), as set forth in Example 3.

The following examples are intended to illustrate embodiments of the invention. They are not intended to limit the scope of the invention in any way.

EXAMPLE 1
Identification and Characterization of MOPs 1-5 cDNAs and Encoded Proteins

We employed an iterative search of human expressed sequence tags to identify novel basic-helix-loop-helix-PAS (bHLH-PAS) proteins that might interact with either the Ah receptor (AHR) or the Ah receptor nuclear translocator (ARNT). In this example, we describe the identification and characterization of five new "Members of the PAS superfamily," or MOPs 1-5, that are similar in size and structural organization to the AHR and ARNT.

METHODS

Search Strategy. The bHLH-PAS domains of the huAHR, huARNT, drSIM, and the PAS domain of drPER were used as query sequences in BLASTN searches of the GenBank database between December of 1994 and October of 1995, using the following default values: Database = NR, expect = 10, word length = 12 (Altschul et al., J. Mol. Biol. 215: 403-410, 1990). Preliminary experiments comparing AHR and PER led us to define candidate ESTs as those "hits" that yielded scores of 150 or higher. As a method to confirm the similarity of these EST sequences to known bHLH-PAS proteins, each candidate EST was subsequently compared to the NR subset of GenBank using the BLASTX program, matrix = blosum 62, word length = 3. Only ESTs that retrieved known bHLH-PAS proteins by this method of confirmation were further characterized.

Oligonucleotide Sequences: Sequences of oligonucleotides are given below. In cases where the oligonucleotide was used in gel shift assays, the 6 bp target sequence is underlined.

OL21 5' CGAGGTCGACGGTATCG 3'
OL22 5' TCTAGAACTAGTGGATC 3'
OL124 5' CCCAAGCTTACGCGTGGTCTTTGAAGTCAACCTCACC 3'
OL145 5' AGCTCGAAATTAACCCTCACTAAAGG 3'
OL176 5' CGGGATCCTTACACATTGGTGTTGGTACAGATGATGTACTC 3'
OL180 5' GCGTCGACTGATGAGCAGCGGCGCCAACATCACC 3'
OL201 5' GATAAGAATGCGGCCGCAGATCTGGGTCCGAAGCACACG 3'
OL202 5' CATTACTTATCTAGAGCTCG 3'
OL226 5' CGGGATCCTCATGGCGGCGACTACTGCCAACC 3'
OL365 5' GACAGTTGCTTGAGTTTCAACC 3'
OL386 5' TTATGAGCTTGCTCATCAGTTGCC 3'
OL387 5' CCTCACACGCAAATAGCTGATGG 3'
OL392 5' CCGCTCGAGTGATGAGCAGCGGCGCCAACATCACC 3'
OL393 5' CCGCTCGAGTGGCAGCTACAGGAATCCACC 3'
OL404 5' GCGGTACCGGGACCGATTCACCATGGAG 3'
OL414 5' TCGAGCTGGGCAGGGTACGTGGCAAGGC 3'
OL415 5' TCGAGCCTTGCCACGTACCCTGCCCAGC 3'
OL418 5' GTAAAACGACGGCCAGT 3'
OL419 5' GGAAACAGCTATGACCATG 3'
OL443 5' TCGAGCTGGGCAGGGTGCGTGGCAAGGC 3'
OL444 5' TCGAGCCTTGCCACGCACCCTGCCCAGC 3'
OL445 5' TCGAGCTGGGCAGGTCACGTGGCAAGGC 3'
OL446 5' TCGAGCCTTGCCACGTGACCTGCCCAGC 3'
OL447 5' TCGAGCTGGGCAGGTTGCGTGGCAAGGC 3'
OL448 5' TCGAGCCTTGCCACGCAACCTGCCCAGC 3'
OL450 5' TACTGGCCACTTACTACCTGACC 3'
OL456 5' AACCAGAGCCATTTTTGAGACT 3'
OL477 5' GCTCTAGAGGCCACAGCGACAATGACAGC 3'
OL479 5' GATCGGAGGTGTTCTATGAGC 3'
OL489 5' TTAGGATGCAGGTAGTCAAACA 3'
OL496 5' GTTCTCCATGGACCAGACTGA 3'
OL499 5' CGGGTACCCTGGGCCCTACGTGCTGTCTC 3'
OL500 5' CGGCTAGCCTCTGGCCTCCCTCTCCTTGATGA 3'
OL514 5' CTGGGAGCCTGCCTGCCTTCA 3'
OL520 5' CCCAAGGAGAGGCGTGAT 3'
OL540 5' GGGATCCTCGTCGCCACTG 3'
OL541 5' ATGCAGTACCCAGACGGATTTC 3'
OL560 5' TGCACGGTCACCAACAGAG 3'
OL561 5' TTGCCAGTCGCATGATGGA 3'
OL565 5' CTGAACAGCCATCCTTAG 3'
OL568 5' AGCTTGCCCTACGTGCTGTCTCAG 3'
OL569 5' AATTCTGAGACAGCACGTAGGGCA 3'
OL590 5' AGAGGTGCTGCCCAGGTAGAA 3'
OL611 5' CAATGATGAGGGAAACACTG 3'
OL657 5' CGGGATCCCGTCAACTGGAGATGAGCAAGGAG 3'
OL665 5' CTGCAAAAATCCGATGACCTCTT 3'
OL681 5' CGGGCAGCAGCGTCTTC 3'
OL682 5' GCGTCCGCAGCCCCAGTTG 3'
OL683 5' TTCAATGTTCTCATCAAAGAGC 3'
OL684 5' GAACAGTTTTATAGATGAATTGGC 3'
OL689 5' GAGGTGTTTCAATTCATCGTCT 3'
OL715 5' GGGATCCGTGACCGATTCACCATGGAG 3'
OL716 5' CTGCAGGTCACACAACGTAATTCACACA 3'
OL717 5' GGGATCCGTATGACAGCTGACAAGGAG 3'
OL718 5' GGTCGACGTCACAGGACGTAGTTGACACA 3'
OL719 5' GAATCCATGAGCAAGGAGGCCGTG 3'

OL720 5' GGTCGACGTCAAACAACAGTGTTAGTTGA 3'
OL721 5' GGGATGCGTATGGATGAAGATGAGAAAGAC 3'
OL722 5' GGTCGACGCTAGACCGAGTGTGTGCA 3'

Cloning strategy. To obtain extended open reading frames for each EST, an anchored-PCR strategy was employed to amplify additional flanking sequence from a variety of commercial cDNA libraries that were constructed in the phagemid Lambda Zap (Tissues; HepG2, Fetal Brain and Skeletal Muscle; Stratagene, La Jolla, CA) (Table 1) (Innis et al. (eds), *PCR Protocols: a Guide to Methods and Applications*, Academic Press, San Francisco, 1990). The resulting PCR products were subjected to agarose gel electrophoresis, transferred to a nylon membrane and analyzed by hybridization with a ^{32}P-labeled probe generated from the corresponding parent EST plasmid (Table 1). After autoradiography, the positive PCR products were purified by gel electrophoresis and cloned using the pGEM-T vector system (Promega, Madison, WI). Dideoxy sequencing was performed to characterize each positive clone.

TABLE 1

MOP cDNA Clone Information

Row 1: clones (in parentheses) containing the candidate ESTs were requested from their laboratory of origin.

Row 2: the Genbank accession number for each original EST is indicated.

Row 3: oligonucleotides used in library screening. Sequence information generated from this original clone was used to design oligonucleotides for use in an anchored PCR strategy, whereby gene-specific and vector-specific primers were used to amplify 5' and 3' portions of the cDNA. Vector specific primers corresponded to modified T3 (5', OL145) or T7 (3', OL146) primers. A matrix of gene-specific primers against annealing temperature (50-65°C) was attempted for each clone, generally leading to at least one successful reaction.

Row 4: the cDNA libraries from which additional sequence of positive clones was identified.

Row 5: size of ORFs. We define a complete ORF by the presence of an in-frame stop codon 5' to a methionine codon that lies within a Kozak consensus sequence for translational initiation. The 3' end of open reading frames are defined by the presence of an in-frame termination codon. An asterisk (*) denotes a clone which does not meet these criteria (see text).

Row 6: Genbank accession numbers of the final MOP cDNAs

are given.

	MOP1/HIF1α	MOP2	MOP3	MOP4	MOP5
Laboratory of origin (clone desig.)	Bell (hbc025)	IMAGE (67043)	IMAGE (23820, 50519)	Liew (F9047, PL420)	IMAGE (42596)
EST Genbank accession number	T10821	T70415	T77200, H17840	R58054	R67292
Gene-specific oligo used in PCR	OL365 (5')	OL456 (5') OL496 (3') OL514 (3') OL541 (3')	OL489 (5')	OL520 (5')	OL540 (5')
Library screened	HepG2	HepG2	Fetal brain	HeLa	HepG2
ORF size (a.a.'s)	826	870	624	642*	412*
Final cDNA Genbank accession number	U29165	U51626	U51627	U51625	U51628

Plasmid Construction for Expression *In Vivo*. Sequence information from each EST was used to design PCR primers for the amplification of cDNA from commercially available libraries. Expression plasmids were constructed by standard protocols (Sambrook et al., 1989). For a summary of clone designations, PCR primers, DNA templates and GenBank accession numbers, refer to Table 1. A brief description follows.

MOP1 expression vectors. Oligonucleotides OL404 and OL365 were used as primers in a PCR to amplify a 970 bp fragment from a HepG2 cell cDNA library. This fragment was cloned into the pGEM-T vector in the T7 orientation and designated PL439. To generate pGMOP1, the *SalI*/*XhoI*

fragment of hbc025 was subcloned into *Sal*I digested PL439. To increase transcription efficiency of the MOP1 cDNA, pGMOP1 was digested with *Kpn*I and *Sac*I and this fragment subcloned into the corresponding sites of pSputk generating PL415 (Stratagene, La Jolla, CA) (Falcone & Andrews, Mol. Cell. Biol. 11: 2656-2664, 1991). The complete ORF of the MOP1 cDNA was amplified using the PCR and oligonucleotides OL425 and OL536. This fragment was digested with *Bam*HI and ligated into the *Bam*HI site in the pSport polylinker (Life Technologies, Inc.). This plasmid was designated PL611.

MOP2 expression vectors. PCR was employed using OL477 and OL450 to amplify a 931 bp MOP2 fragment from a HepG2 cDNA library. This fragment was cloned into pGEM-T in the SP6 orientation and designated PL424. Using OL560 and OL590, PCR amplification from this same library yielded a 3' fragment of the MOP2 cDNA. This fragment was cloned into pGEM-T in the SP6 orientation and was designated PL445. PL424 was digested *Sal*I and *Eco*RI and the fragment ligated into a *Sal*I/*Eco*RI digested PL445 to generate a full ORF MOP2 expression vector designated PL447. The complete ORF of the MOP2 cDNA was cloned into pSport as follows; PL447 was digested with *Sac*II, treated with the Klenow fragment of DNA Polymerase I in the presence of dNTPs, and subsequently digested with *Sal*I. This fragment was purified and ligated into pSport digested with HindIII, repaired with Klenow, then digested with *Sal*I. This construct was designated PL477.

MOP3 expression vectors. Using the primers OL145 and OL489 and a human fetal brain cDNA library as template, the PCR was used to obtain a 1380 bp fragment. This fragment was isolated and cloned into pGEM-T as above, and this plasmid designated PL487. A fragment of MOP3 was obtained by the PCR using Pfu polymerase (Stratagene), primers OL657 and OL689 and PL487 as template. To obtain a full length MOP3 cDNA fragment, the megaprimer fragment obtained above was used in the PCR

against oligonucleotide OL611 using IMAGE clone 50519 as a template (Sarkar & Somers, Biotechniques 8: 404-407, 1990). This product was cloned into pGEM-T in the SP6 orientation as above and designated PL425.

MOP4 expression vectors. Using primers OL520 and OL145 and a HepG2 cDNA library as template, the PCR was performed to isolate a 5' fragment of the MOP4 cDNA. This fragment was cloned in the T7 orientation of pGEM-T and designated PL448. The cDNA insert of the phage clone F9047 (C.C. Liew, Toronto, CA) was amplified by the PCR using oligonucleotides OL418 and OL419 and subcloned into the pGEM-T vector (Hwang et al., J. Mol. Cell. Cardiol. 26: 1329-1333, 1994). This clone was designated PL420. An *Eco*RI fragment of PL448 was isolated and cloned into a partially *Eco*RI digested PL420. This clone was subjected to the PCR with oligonucleotides OL698 and OL146, and the fragment cloned into the pGEM-T vector, and designated PL545.

MOP5 expression vectors. The PCR was used to obtain a 1260-bp fragment of the MOP5 gene using oligonucleotides OL685 and OL686 and IMAGE clone 42596 as template. This fragment was purified and subcloned into the pGEM-T vector as above in the SP6 orientation. This plasmid was designated PL528 and subsequently digested with *Sal*I and partially digested with *Nco*I. This fragment was ligated into Nco I/*Sal*I-cut pSputk, and the resulting vector designated PL554.

Hypoxia responsive luciferase reporters. The plasmid pGL2EPOEN was constructed as follows: The hypoxia responsive enhancer from the 3' region of the EPO gene was amplified by PCR using oligonucleotides OL499 and OL500 and human genomic DNA as template (amplified fragment corresponds to nucleotides 127 to 321 as reported in the EPO structural gene sequence found in GenBank Accession GBL16588). This fragment was digested with *Kpn*I and *Nhe*I and cloned into the corresponding sites of the plasmid pGL2-Promoter (Promega).

Antibody Production. Antisera against MOP1, MOP2, AHR and ARNT were prepared in rabbits using immunization protocols that have been described previously (Poland et al., Mol. Pharmacol. 39: 20-26, 1991; Pollenz et al., Mol. Pharmacol. 45: 428-438, 1994). Crude antisera was chosen for use in all coimmunoprecipitation experiments and the PI sera from the same rabbit served to preclear the samples. For MOP1, the plasmid hbc025 was digested with *Eco*RI and the 604 bp fragment was treated with the Klenow fragment of DNA polymerase-1 in the presence of dNTPs and cloned into the *Sma*I site of the histidine tag fusion vector pQE-32 (Qiagen, Chatsworth, CA). This clone, designated PL377, was transformed by electroporation into M15(REP4) cells for expression under IPTG induction. The expressed protein was purified from 8 M urea using Ni-NTA agarose, extensively dialyzed against 25 mM MOPS, pH 7.4, 100 mM KCl, and 10% glycerol before its use as an immunogen. For AHR, the human cDNA clone PL71 (Dolwick et al., Mol. Pharmacol. 44: 911-917, 1993) was digested with *Bam*HI and cloned into the corresponding site of the histidine fusion vector pQE31 (Qiagen). The AHR protein fragment was expressed and purified exactly as described for MOP1 (above). Antiserum produced against this protein was designated R2891. For MOP2 a *Sac*I/*Pst*I fragment of PL445 was cloned into *Sac*I/*Pst*I cut pQE-31 to generate PL456. This clone, designated PL456, was transformed into M15(REP4) cells and the protein expressed under IPTG induction. The histidine tagged fusion protein was first extracted in guanidine hydrochloride, dialyzed extensively and purified on Ni-NTA agarose as above. Antiserum produced against this protein was designated R4064. ARNT-specific antisera was raised against huARNT protein purified from baculovirus as previously described (Chan et al., J. Biol. Chem. 269: 26464-26471, 1994).

Northern Protocol. Multiple tissue northern blots containing 2 μg of poly(A)+ mRNA prepared from human heart, brain, placenta, lung, liver, skeletal muscle,

kidney, and pancreas were probed with random primed cDNA fragments using an aqueous hybridization protocol (Clontech, Palo Alto, CA). Hybridization solution contained 5 X SSPE (0.75 M NaCl, 50 mM NaH_2PO_4, 5 mM Na_2EDTA, pH 7.4) 2 X Denhardt's solution (0.04 % w/v Ficoll 400, 0.04% w/v polyvinylpyrrolidone, 0.04 % w/v Bovine Serum Albumin), 0.5 % SDS, and 100 ug/mL heat denatured salmon sperm DNA. A blot was prehybridized for 3-6 hours at 65°C, the hybridization solution was changed and $1-5x10^6$ cpm/mL of a random primed cDNA fragment was added. Samples were hybridized overnight at 65°C, washed twice with 2 X SSC (0.3 M NaCl, 30 mM Na_3Citrate, pH 7.0), 0.5 % SDS at room temperature, once with 1X SSC, 0.1% SDS at the hybridization temperature, and once with 0.1 X SSC, 0.1% SDS at the hybridization temperature.

Yeast Two-Hybrid Analysis. A modified yeast interaction trap was employed to identify those MOPs that could interact with the AHR or ARNT. LexAMOP chimeras were constructed to fuse the bHLH-PAS domains of the MOP proteins with the DNA binding domain of the bacterial protein LexA (amino acids 1-202) (Bartel et al., BioTechniques 14: 920-924, 1993). To amplify the region corresponding to the bHLH-PAS domains of MOP1, OL715 and OL716 were employed in the PCR using PL415 as template. To amplify the region corresponding to the bHLH-PAS domains of MOP2, OL717 and OL718 were employed in the PCR using PL447 as template. To amplify the region corresponding to the bHLH-PAS domains of MOP3, OL719 and OL720 were employed in the PCR using PL486 as template. To amplify the region corresponding to the bHLH-PAS domains of MOP4, OL721 and OL722 were employed in the PCR using PL545. Since a more detailed domain map existed for the AHR, a construct was made with a fine deletion of the transactivation domain. The N-terminal portion of the AHR was amplified by the PCR using oligonucleotides OL180 and OL124 and pmuAHR as template (Dolwick et al., Proc. Natl. Acad. Sci. USA 90: 8566-8570, 1993). This product was

digested with *Kpn*I and *Sal*I, and cloned into the corresponding sites of pSG424 (Sadowski & Ptashne, Nucl. Acids Res. 17: 7539, 1989). This clone was designated PL187. The 3' end of the AHR cDNA was amplified by PCR using oligonucleotides OL201 and OL202 and pmuAHR as template. This product was digested with *Not*I, and cloned into the corresponding site of pSGAhN-delta-166 (Dolwick et al, Proc. Natl. Acad. Sci. USA 90: 8566-8570, 1993). This clone was designated PL188. PL188 was digested with *Kpn*I and *Xba*I, and this fragment cloned into the corresponding sites of PL187. This clone was designated PL204. A cDNA fragment of the AHR was amplified by the PCR using OL392 and OL393 and PL204 as template. This product was cloned using the pGEM-T system, and designated pGTAHR-delta-TAD. This construct was digested with *Xho*I and this fragment ligated into *Sal*I cut pBTM116 (Vojtek et al., Cell 74: 205-214, 1993). This construct was designated pBTMAHR. LexAARNT was constructed by PCR using oligonucleotides OL226 and OL176 and PL87 as template. The PCR product was cloned into pGEM-T as above, and the *Bam*HI fragment cloned into a *Bam*HI digested pGBT9 vector (Clontech). This construct was cut with *Bam*HI and subcloned into a *Bam*HI digested pBTM116. This construct was designated LexAARNT. Following amplification these products were purified and cloned into the pGEM-T vector. These clones were designated PL537 (MOP1), PL538 (MOP2), PL539 (MO3), and PL540 (MOP4). These plasmids were digested with *Bam*HI/*Pst*I (PL537), *Bam*HI/*Sal*I (PL538 and PL540), and *Eco*RI/*Sal*I (PL539), and these fragments ligated into the appropriately digested pBTM116. These clones were designated LexAMOP1, LexAMOP2, LexAMOP3, and LexAMOP4, respectively. Full length expression plasmids harboring the AHR and ARNT were constructed as follows: PL104 (pSporthuAHR) was cut with *Sma*I, the insert purified and subcloned into *Sma*I site of pCW10, and this plasmid was designated PL317. This clone was digested with *Sma*I and subcloned into a *Sma*I cut pRS305, and this clone

designated pRSAHR. The ARNT cDNA (PL101) was digested with *Not* I and *Xho*I, and cloned into the corresponding sites of pSGBMX1. This plasmid was designated PL371, and subsequently was digested with *Not*I and *Xho*I, this fragment cloned into the corresponding sites of pSGBCU11. This clone was designated PL574. The LexAMOP fusion protein constructs were cotransformed with a yeast expression vector containing the full length AHR or ARNT into L40, a yeast strain containing integrated *lacZ* and *HIS3* reporter genes. As controls, LexAAHR and LexAARNT were cotransformed with AHR and ARNT. The strength of interaction was visually characterized by X-Gal (5-bromo-4-chloro-3-indolyl-(-D-galactoside) plate assays, performed after three days growth on selective media (Bohen et al, Proc. Natl. Acad. Sci. USA 90: 11424-11428, 1993). To provide quantitation of the interaction strength, multiple colonies from yeast harboring each bHLH-PAS combination colonies were grown overnight in liquid media. Liquid cultures were grown for 5 hours, and assayed for *lacZ* activity using the Galacto-Light chemiluminescence reporter system (Tropix, Bedford, MA). To determine the effect of AHR agonists on these interactions, yeast were also grown on plates or in liquid culture with and without 1 μM βNF (Carver et al., J. Biol. Chem. 269: 30109-30112, 1994).

To ensure expression of each bHLH-PAS construct, western blot analysis was performed using antibodies raised against the LexA DNA binding domain. Yeast extracts were prepared from 15 mL overnight cultures derived from multiple colonies of yeast expressing each LexAMOP fusion protein. Cultures were subjected to centrifugation at 1200 x g for 5 minutes and the pellet was resuspended in 500 μL of 6 M Guanidinium-HCl, 0.1 M Na-Phosphate Buffer, 0.01 M Tris pH 8.0. This suspension was transferred to a fresh eppendorf tube containing 500 μL of acid washed glass beads (Sigma), and mixed on the max setting in a Bead-Beater (BioSpec, Bartlesville, OK)

for 3 minutes at 4°C. The samples were cleared by centrifugation at 14,000 x g, and 400 μL of supernatant was precipitated with 400 μL of 10 % TCA on ice. After clearing by centrifugation at 14,000 x g for 20 minutes at 4°C, the extracts were resuspended in SDS loading buffer and subjected to SDS-PAGE analysis. Following electrophoresis, proteins were transferred to nitrocellulose membrane and detected with LexA antisera and secondary antibodies linked to alkaline phosphatase by standard protocols (Jain et al., J. Biol. Chem 269: 31518-31524, 1994).

Transient transfection of Hep 3b cells. pGL2EPOEN was cotransfected with pSport, PL464 (pSportMOP1), or PL477 (pSportMOP2) using the Lipofectin protocol (Life Technologies, Inc.). Briefly, the expression vector was mixed with the epo-reporter and the beta-galactosidase control plasmid pCH110 (Clontech) at a 3:1 charge ratio of TFX-50 reagent (Promega) and incubated for 15 minutes at room temperature. The lipofection media (200 μL) was added to Hep3b cells in 4 cm^2 plates in the presence of serum. The cells were incubated at 37°C for 2 hr. Following incubation, fresh media was added, and the cells were incubated for an additional 48 hours prior to harvesting. Cell extraction and beta-galactosidase assays were performed using the Galacto-Light assay according to manufacturer's protocols (Tropix).

Coimmunoprecipitation with Hsp90. Each MOP construct was *in vitro* translated in the presence of [^{35}S]-methionine in a TNT coupled transcription/translation system (Promega). Hsp90 immunoprecipitation assays were performed with monoclonal antibody 3G3p90 or a control IgM antibody, TEPC 183 (Sigma) essentially as described. Each immunoprecipitation was subjected to SDS-PAGE, and the resulting gel was dried. The level of radioactivity in each coprecipitated protein band was quantified on a Bio-Rad GS-363 Molecular Imager System. The amount of protein immunoprecipitated with the Hsp90 antibody is

presented as a percentage of the amount of murine AHR immunoprecipitated in parallel assays.

RESULTS

EST Search. Our initial BLAST searches in December 1994 were performed with the bHLH-PAS or PAS domains of all family members known at that time (AHR, ARNT, SIM and PER). In these searches we identified an EST clone, hbc025, derived from human pancreatic islets (Table 1) (Takeda et al., Hum. Mol. Genet. 2: 1793-1798, 1993). To confirm this similarity, we performed a BLASTX search, comparing hbc025 to the GenBank database and found that this sequence was most homologous to Drosophila SIM. This EST clone was designated MOP1. The bHLH-PAS domains of all family members, including MOP1, were again searched from May to October of 1995. Human ESTs that recorded BLASTN scores above 150 were again retrieved and confirmed using the BLASTX algorithm. This routine resulted in the discovery of six ESTs with significant homology to the bHLH-PAS domains of known members (Table 1).

cDNA Cloning. In order to more completely characterize the similarities and domain structures of the candidate clones, an anchored-PCR strategy was employed to obtain additional flanking cDNA sequence using phagemid libraries as a template. Comparison of amino acid sequences of these bHLH-PAS proteins is displayed in Figure 2. Upon characterization of the open reading frames, it was learned that two of these ESTs (F06906 and T77200) corresponded to the same gene product (Table 1). Thus, we designated these remaining five unique cDNAs as "Members of PAS superfamily" or MOPs 1-5. The PCR strategy provided what appeared to be the complete ORFs of MOP1, MOP2 and MOP3 based upon the following criteria: (1) at their 5' ends these clones contain an initiation methionine codon (AUG) downstream of an in-frame stop codon, and (2) at their 3' ends these clones contain an in-frame stop codon followed by no obvious open reading

frames. In addition, the nucleotide sequences flanking of the MOP1 and MOP2 most 5' AUG codons (see GenBank accessions U29165 and U51626) are in reasonable agreement with the proposed optimal context for translational initiation, i.e., CCACCAUGG (Kozak, Cell 44: 283-292, 1986; Kozak, Nucl. Acids Res. 15: 8125-8132, 1987).

Using the same anchored-PCR technique, we were unable to obtain the complete open reading frames of MOP4 or MOP5. This may have been due to the low copy number of these mRNAs in the tissues from which our PCR source cDNA was obtained (see below). We did identify a potential start methionine for MOP4 and the 3' stop codon for MOP5 (Figure 2). Our preliminary designation of the MOP4 start methionine is tentative and is based only on its proximity to the start methionines of MOP1, MOP2, MOP3, AHR and SIM (Figure 2) (Burbach et al., Proc. Natl. Acad. Sci. USA 89: 8185-8189, 1992; Nambu et al., Cell 67: 1157-1167, 1991). The fact that only one of the six nucleotides flanking the MOP4 AUG codon (ATTTAATGG) matches the consensus sequences for optimal translational initiation provides an indication that a more 5' initiation codon may exist. Therefore, the initiation codon of MOP4 is uncertain and that of MOP5 remains to be identified. This low level expression is consistent with our difficulties in amplifying these cDNAs by PCR (see above) and suggests that expression may be limited to specific cell types or developmental time periods not identified in our study.

Tissue Specific Expression. To characterize the tissue specific expression patterns of the MOP mRNAs, Northern blots of poly A(+) RNA from eight human tissues were probed with random primed cDNA restriction fragments. Single transcripts of 3.6 kb (MOP1/HIF1α), 6.6 kb (MOP2) and 3.2 kb (MOP3) were detected. Expression levels of each mRNA varied significantly between tissues, with MOP1 being highest in kidney and heart, MOP2 highly expressed in placenta, lung, and heart, and MOP3 highly expressed in skeletal muscle and brain. No detectable message was

detected for MOP4 or MOP5 by our northern blot protocol.

Identification of Novel AHR or ARNT Partners:

1. Interaction of MOPs with the AHR and ARNT *in vitro*; Coimmunoprecipitation experiments. We first performed coimmunoprecipitation experiments to determine if MOPs 1-4 had the capacity to interact with either the AHR or ARNT *in vitro*. These proteins were expressed in a reticulocyte lysate system in the presence of ^{35}S-methionine and then incubated in the presence or absence of the AHR or ARNT. Complex formation was assayed by coimmunoprecipitation with AHR or ARNT specific antisera, followed by quantitation of coimmunoprecipitated ^{35}S-labeled MOP by phosphoimage analysis. Interactions were identified by a reproducible increase in an AHR or ARNT-dependent precipitation of MOP protein. Because we have observed considerable variability in this coimmunoprecipitation assay, each experiment was performed at least three times.

In the AHR interaction studies, we observed that MOP3 was coimmunoprecipitated with AHR. The positive control, ARNT-AHR interaction, was also reproducible, but weaker. Neither MOP1, MOP2 or MOP4 could be shown to interact with the AHR by this protocol. The ARNT protein displayed a broad range of interactions and was shown to coimmunoprecipitate with AHR (positive control), MOP1 and MOP2, but not MOP3 or MOP4.

2. Interaction of MOPs with the AHR and ARNT *in vivo*; Yeast two-hybrid experiments. To determine if MOP-AHR or MOP-ARNT complexes could form *in vivo*, a modified interaction trap was employed (Fields & Song, Nature 340: 245-246, 1989; Chien et al., Proc. Natl. Acad. Sci. USA 88: 9578-9582, 1991) (Figure 3). Fusion proteins were constructed in which the DNA binding domain of the bacterial repressor, LexA, was fused to the bHLH-PAS domains of the MOP proteins (Fig. 3A). The bHLH-PAS domains were chosen because they harbor both the primary and secondary dimerization surfaces of this family of proteins and they do not harbor transcriptional activity

that would interfere with this assay (Jain et al., 1994, *supra*). Interactions were tested by cotransformation of each LexAMOP construct with either the full length AHR or ARNT into the L40 yeast strain, which harbors an integrated *lacZ* reporter gene driven by multiple LexA operator sites. In this system, LexAMOP fusions which interact with AHR or ARNT drive expression of the *lacZ* reporter gene.

We assessed the relative strength of these interactions by both a direct *lacZ* plate assay and by quantitation of the reporter activity in a liquid culture (Figure 3). In all cases, these two methods of detection were equivalent. To test the validity of this model system as a method to detect bHLH-PAS interactions, LexAAHR and LexAARNT constructs were cotransformed with either the full length ARNT or AHR. In these control experiments, we were able to demonstrate the specificity of AHR-ARNT interaction and its dependence on the presence of the agonist βNF. The LexAAHR-ARNT interaction in the presence of βNF was 913 fold above background, while the LexAARNT-AHR interaction in the presence of βNF was 14 fold above background. Both combinations showed ligand inducibility. The LexAAHR-ARNT interaction in the presence of βNF was 6.4 fold over LexAAHR-ARNT in the absence of ligand, while the LexAARNT-AHR interaction in the presence of βNF was 2.0 fold over LexAARNT-AHR in absence of ligand. Despite our ability to readily detect the agonist-induced LexAARNT-AHR interaction in the two hybrid system, we were unable to detect any LexAMOP that could interact with the AHR. That is, none of the LexAMOP fusion proteins appeared to interact with cotransformed AHR and drive *lacZ* expression in the absence or presence of ligand.

Two of four MOP proteins tested were found to interact with ARNT in the two hybrid assay. Both the LexAMOP1 and LexAMOP2 interactions with full length ARNT were extremely robust, 36 and 28 fold above background,

respectively (Fig. 3B). When compared with the LexAAHR-ARNT interaction in the presence of βNF, the LexAMOP1-ARNT and LexAMOP2-ARNT interactions were 24% and 69% as intense. These differences in LexAMOP1-ARNT and LexAMOP2-ARNT interaction could be attributed to differences in expression levels or to subtle differences in vector construction. To control for relative expression of the LexAMOP fusions, protein extracts were prepared and western blot analysis was performed with LexA specific antisera. We observed expression for each LexAMOP fusion proteins, indicating that negative results with LexAMOP3 and LexAMOP4 are not due to lack of expression.

3. DNA binding and specificity. Prompted by the observation that MOP1 and ARNT and MOP2 and ARNT specifically interact, we next examined the ability of MOP-ARNT dimeric complexes to bind those DNA response elements recognized by other bHLH-PAS protein complexes. Reports from a number of laboratories have demonstrated that bHLH-PAS dimers can bind to a variety of DNA elements: "DRE," TNGCGTG (Denison et al., J. Biol. Chem. 264: 16478-16482, 1989); "CME," ACGTG (Wharton et al., Development 120: 3563-3569, 1994); "SAE," GTGCGTG (Swanson et al., J. Biol. Chem. 270: 26292-26302, 1995); and "E-box," CANNTG (Sogawa et al., Proc. Natl. Acad. Sci. USA 92: 1936-1940, 1995; Swanson et al., 1995, *supra*). Using a gel-shift assay, we observed that MOP1-ARNT complexes specifically bound CACGTG and TACGTG, while the complex failed to bind GTGCGTG, TTGCGTG, and a non-palindromic E-box, CATGTG. Previous reports have demonstrated that ARNT homodimers are capable of binding the CACGTG sequence *in vitro*, and that this complex can drive reporter gene expression *in vivo* (Sogawa et al., 1995 *supra*, Swanson et al., 1995, *supra*). Our results suggest that the MOP1-ARNT dimeric complex binds the CACGTG oligonucleotide with a higher affinity than either MOP1 or ARNT alone. MOP1 failed to form a productive DNA binding complex with the

AHR with any of the bHLH-PAS family response elements. As a comparison of MOP1-ARNT and MOP2-ARNT DNA binding, we provide results from gel shift assays using a double-stranded oligonucleotide containing a core TACGTG hexad binding site. WE observed that both MOP1-ARNT and MOP2-ARNT bound the TACGTG-containing oligonucleotide with approximately equal capacity and neither ARNT, nor MOP1, nor MOP2 could bind this DNA sequence alone. As additional controls, we confirmed the presence of the MOP proteins in the complex by showing that antisera raised against these proteins retarded the mobility of the complex.

4. Interaction of MOPs with Hsp90. In an effort to assess a MOP's ability to interact with Hsp90, we performed coimmunoprecipitation assays with anti Hsp90 antibodies. Given the remarkable stability of the Hsp90 complex with the AHR from the C57BL/6J mouse, we used this receptor species as a reference and compared all interactions relative to it. As additional controls, we immunoprecipitated ARNT and the human AHR as negative and positive controls, respectively. Despite our ability to readily detect huAHR-Hsp90 interactions, we were unable to detect ARNT, MOP2 or MOP5 interactions with Hsp90. In contrast, huAHR, MOP1, MOP3 and MOP4 all immunoprecipitated with HSP90-specific antisera. MOP3 formed the tightest interaction with HSP90, followed by the huAHR, MOP4 and MOP1 (71%, 53%, 31% and 17%, respectively).

DISCUSSION

Since cDNAs encoding the complete open reading frames for MOPs 1-3 were available, most of the studies described in this example focused on those proteins. MOP4 was also included in some studies since our clone contained the sequences involved in dimerization, transcriptional activation and DNA binding of other bHLH-PAS proteins. Given the limited sequence information

on MOP5, this clone was typically not included in functional studies.

Tissue specific expression. We observed that each MOP mRNA displayed a unique tissue specific distribution with MOP1 being highest in kidney and heart, MOP2 highly expressed in placenta, lung, and heart, and MOP3 highly expressed in skeletal muscle and brain. Previous studies conducted in our laboratory indicated that ARNT is expressed highly in skeletal muscle and placenta, while the AHR is most prevalent in placenta, lung, and heart (Carver et al., 1994, *supra*; Dolwick et al., 1993, *supra*). The observation that these bHLH-PAS proteins are coexpressed in a variety of tissues supports the idea that cross talk between these signaling pathways may be occurring *in vivo* and that multiple tissue specific interactions may be taking place. We also observed that AHR and MOP2 have very similar expression profiles in human tissues. An additional and equally important interpretation of these unique MOP expression profiles is that unidentified partners exist for these bHLH-PAS proteins and that they regulate a number of undescribed biological pathways.

Interactions. Our interaction screening strategy was based on the large amount of functional data and the detailed domain maps available for the AHR and ARNT. An important assumption used in the design and interpretation of our studies is that some of the MOPs may be constitutive interactors in vivo (like ARNT) and others may be conditional interactors that require activation in order to dimerize in vivo (like AHR). We chose to employ coimmunoprecipitation as an initial interaction screen for a number of reasons. First, AHR and ARNT-specific antibodies are available that have been shown to pull down AHR-ARNT complexes. This suggests that if MOP-AHR or MOP-ARNT interactions occurred *in vitro*, that these same antibodies would recognize and pull down such complexes. Second, data from a number of laboratories, using

independently derived antibodies indicates that coimmunoprecipitation of AHR-ARNT complexes is independent of AHR-ligand. This observation suggests that AHR or ARNT interactions with conditional MOP proteins might still be identified by coimmunoprecipitation even in the absence knowledge about how to activate a conditional MOP (e.g., identification of a ligand).

As a secondary screen to characterize interacting MOPs, we employed a yeast interaction trap commonly referred to as the "Two Hybrid Assay". Support for use of this system comes from our previous observation that LexA-AHR chimeras are functional in yeast and provide a good model of AHR signaling and ARNT interaction. In addition, this method provides an independent confirmation of those interactions identified by coimmunoprecipitation and also provides a demonstration that interactions can occur in vivo. One major limitation of this system is that it may be insensitive at detecting conditional MOPs that require activation prior to dimerization. An example of this can be seen with the AHR and ARNT. In the absence of ligand, the AHR appears to reside primarily in the cytosol and ARNT appears to be primarily nuclear. This compartmentalization appears to be part of a cellular mechanism to prevent interaction of these proteins and minimize constitutive activity of the complex. It is important to point out that compartmentalization is only one component of AHR regulation, since ligand dependent DNA binding does occur in vitro in the presence of ARNT. Nevertheless, *in vivo* systems like the Two Hybrid Assay may yield false negative results for conditional MOP protein interactions that require an upstream activation event prior to nuclear transloction.

In light of the above considerations, our interpretation of the coimmunoprecipitation and two hybrid interaction results were as follows: First, since the MOP1-ARNT and the MOP2-ARNT interactions were confirmed in two independent systems these interactions should be

pursued further (see below). Second, the observation that MOP3 interact with the AHR *in vitro*, but not *in vivo*, suggests that MOP3 may be a conditional MOP that has the capacity to interact with the AHR *in vivo*. This idea gained support from Hsp90 interaction studies (below). Third, the suspicion that MOP3 is a conditional bHLH-PAS protein, coupled with the observation that MOP3 and AHR have the disparate expression profiles led us to delay study of this interaction until we learn how to activate MOP3 or until we have evidence that these two proteins are expressed in the same cell type. Fourth, our observation that ARNT can form dimers with two out of four MOPs examined suggests that ARNT is a highly promiscuous bHLH-PAS partner that may be a focus of cross talk between different MOP signaling pathways. The multiplicity of ARNT partnerships is supported by recent observations from a number of laboratories (Sogawa et al, 1995, *supra*; Swanson et al., 1995, *supra*).

MOP1 and MOP2 interactions with ARNT. The concordance of the coimmunoprecipitation and two hybrid data led us to pursue the MOP1-ARNT and MOP2-ARNT interactions further. Given the pairing rules deduced from the interaction studies described above, we next attempted to determine if the MOP1-ARNT and MOP2-ARNT complexes bound specific DNA sequences *in vitro*. Earlier reports indicated that the basic region of each bHLH partner generates specificity for a distinct DNA half-site of at least 3 bp. Data from a number of laboratories has indicated that the ARNT protein displays specificity for the 3' GTG half site of the hexad target sequence, 5'NNCGTG3', where 5'NNC is the half site of the ARNT partner. To determine the half site specificity of the MOP1 protein when complexed with ARNT, we used gel shift analysis with oligonucleotides containing substitutions at the two variable 5' positions of 5'NNCGTG3'. These preliminary experiments indicated that MOP1-ARNT complex had greatest affinity for the 5'CAC and 5'TAC half sites.

Because the MOP1 and MOP2 basic regions differed by only one amino acid residue and since this residue is not thought to be in a DNA contact position, we hypothesized that MOP2 would bind the same DNA sequences. To confirm this, we performed MOP2-ARNT gel shift assays using a double stranded oligonucleotide containing a core TACGTG hexad binding site. We observed that both MOP1-ARNT and MOP2-ARNT bound the TACGTG containing oligonucleotide, that neither MOP1 nor MOP2 could bind this sequence in the absence of ARNT. As additional controls, we confirmed the presence of the MOP1 and MOP2 proteins in the complex by showing that antisera raised against these proteins retarded the mobility of the complex.

To assay MOP1-ARNT and MOP2-ARNT interactions *in vivo*, we constructed a luciferase reporter driven by the hypoxia responsive TACGTG containing enhancer from the human EPO gene. Our transient expression experiments in Hep3B cells consisted of cotransfection of this reporter with vector control, MOP1, or MOP2 in the presence or absence of cobalt chloride to stimulate the hypoxia heme sensor. ARNT has been shown previously to be expressed in Hep3B cells. This experiment confirmed that the TACGTG-containing enhancer sequence is responsive to cobalt and cotransfected MOP1 or MOP2 under normal oxygen tension. The transfected MOP1 construct appeared to be responsive to hypoxia (3.5-fold over control), while the MOP2 construct was only slightly responsive (1.2-fold). MOP2 was more potent than MOP1 in driving expression of this reporter gene both in the presence and absence of cobalt chloride. This difference in efficacy of the MOP1 and MOP2 reporter plasmid in Hep3B cells could be explained by three possibilities: (1) the relative potency of the MOP2 transactivation domain may be much greater than MOP1; (2) the relative expression of MOP2 may be greater in this transient expression system than MOP1; or (3) the MOP1 may be partially repressed *in vivo*, by HSP90, while MOP2 is

not (see HSP90 discussion below). Given that our MOP2 antisera are not useful in western blots, we could not assess the relative expression or stability of the MOP1 and MOP2 clones in this system.

MOP3 is a conditionally active bHLH-PAS protein. Data from a number of laboratories suggests that Hsp90 represses AHR activity by anchoring the receptor in the cytosol away from its nuclear dimeric partner ARNT. Ligand binding appears to weaken the Hsp90 association and induce a translocation of the Hsp90-AHR complex to the nucleus where dimerization with ARNT can occur.

Two lines of evidence suggest that MOP3, like the AHR, may be a conditionally active bHLH-PAS protein and that in the absence of an unidentified cognate ligand, might be repressed and unable to dimerize *in vivo*. First, MOP3 interacts with HSP90 even more efficiently than human AHR, suggesting that MOP3 may be functionally repressed or anchored in the cytosol like the AHR. Second, MOP3 interacts with AHR in the coimmunoprecipitation assay, but not in the yeast interaction trap. Similarly, the AHR interacts with ARNT in the coimmunoprecipitation assay, but interacts weakly, if at all, in the absence of ligand activation.

Alternative explanations for the different MOP3-AHR interaction results obtained from our *in vitro* and *in vivo* systems should also be considered. For example, the structure of MOP3 may be different than the AHR and ARNT, such that positioning of the LexA domain adjacent to the bHLH-PAS domain may sterically hinder dimerization surfaces within this protein or lead to improper subcellular localization or instability of the chimera. One example of the potential negative impact of context sensitivity in the two-hybrid system can be observed in Figure 3. The LexAAHR-ARNT interaction is 14.7 times more robust than the LexAARNT-AHR interaction. In addition, the LexAAHR-ARNT interaction is more responsive to the AHR ligand βNF than the LexAARNT-AHR combination (6.4-fold and

2.0-fold, respectively). This difference cannot be explained by the relative transactivation potencies of the transactivation domains of AHR and ARNT in yeast, and therefore must be the result of context sensitivity. A final consideration is that coimmunoprecipitations may be capable of detecting weak interactions that cannot be maintained at the low cellular concentrations of the various MOPs. Thus, the MOP3-AHR dimerization may be too weak to occur *in vivo*. In this regard, we have previously reported ARNT-ARNT homodimers that bind specific DNA enhancer sequences *in vitro*, but they are weakly active, if active at all, *in vivo* (Swanson et al., 1995, *supra*).

It is also important to note that MOP1 and MOP4 also interact with HSP90 in the coimmunoprecipitation assay, albeit less strongly than MOP3 or human AHR. The relatively weak interaction of MOP1 with HSP90 may be an indication that this protein is partially repressed *in vivo* and that it may have both constitutive and conditional activity. Such a phenomenon might explain why MOP1 has less transcriptional activity in our *in vivo* systems than MOP2, which does not interact with HSP90. Finally, MOP4 did not interact with the AHR or ARNT in either the coimmunoprecipitation assay or the interaction trap. Although our experience with AHR indicates that interactions with conditional bHLH-PAS proteins can be observed by coimmunoprecipitation assays, MOP4's interaction with HSP90 may also indicate a requirement for activation and may inhibit the sensitivity of detecting interactions *in vivo*.

EXAMPLE 2
MOP3 Forms Transcriptionally Active Complexes with Circadian and Hypoxia Factors

As described above, a number of "orphan" bHLH-PAS proteins have emerged from searches of expressed sequence tag databases and low stringency hybridization screens. For newly discovered bHLH-PAS proteins that have close homologs (e.g., HIF1α and HIF2α (MOP2), or ARNT and

ARNT2), partnering and DNA binding specificity can often be predicted from amino acid sequence similarities in their bHLH-PAS domains. For divergent orphans like MOP3, MOP4, and MOP5, amino acid sequence does not provide the information necessary for similar predictions. To characterize this class of orphans, we have employed a series of assays that allow us to: (i) identify heterodimeric partnerships, (ii) determine the DNA response element bound by these heterodimers, (iii) verify that these complexes drive transcription in mammalian cells, and (iv) identify those tissues where these partnerships may occur. This example describes application of this approach to two bHLH-PAS orphans, MOP3 and MOP4.

MATERIALS AND METHODS

Reagents. Oligonucleotides were supplied by GIBCO/BRL and designated as follows:

OL522 5'-GACAGTATCACGCCTCTCCTT-3'
OL579 5'-AGCGGCGTCGGGATAAAATGA-3'
OL595 5'-ATGCTGAACTGTGCCGAAAACTGT-3'
OL656 5'-GAACAGTGGGGTGGGTCCTCTTT-3'
OL990 5'-GGAATTCTGAGTCTGAAC-3'
OL991 5'-GGAATTCCACGCTCAGG-3'
OL992 5'-GGAATTCTGAGTCTGAAC$(N)_{13}$CCTGAGCGTGGATTCC-3'
OL1116 5'-GATCGGACACGTGACCATTGGTCACGTGTCCATTGGACACGTGACC-3'
OL1117 5'-GATCGGTCACGTGTCCAATGGACACGTGACCAATGGTCACGTGTCC-3'
OL1155 5'-GATCGGATACGTGACCATTGGTTACGTGTCCATTGGATACGTGACC-3'
OL1156 5'-GATCGOTCACGTATCCAATGGACACGTAACCAATGGTCACGTATCC-3'.

The yeast LexA fusion plasmid pBTM116 was provided by P. Bartel and S. Fields (State University of New York, Stony Brook). The yeast strain L40 was a kind gift of S. Hollenberg (Fred Hutchinson Cancer Research Center, Seattle, WA). The yeast strain AMR70 was constructed by Rolf Sternglanz, and was a kind gift of S. Hollenberg. LexA antiserum was a kind gift of J. W. Little (University of Arizona). pSGBCU11 was a kind gift of Stephen Goff (CIBA-Geigy, Research Triangle Park, NC). Human CLOCK was a kind gift of T. Nagase (Kazusa DNA Research Institute, Chiba, Japan). Mammalian expression

vectors were purchased from GIBCO/BRL (pSVSport) and Promega (pTarget). Antibodies specific for MOP3 and MOP4 were prepared against peptides specific for each protein as described (Poland et al., 1991, Mol. Pharmacol. 39:20-26). The MOP3 peptide sequence was DNDQGSSSPSNDEAAC, and the MOP4 peptide sequence was KDKGSSLEPRQHFNALDVGC.

Expression Plasmid Construction. Yeast expression plasmids harboring the LexA DNA binding domain fused to the bHLH-PAS domains of HIF1α (PL856), HIF2α (PL857), MOP4 (PL859), AHR (PL739), and ARNT (PL701) have been described (Example 1). LexAbHLH-PAS fusion plasmids for MOP3 (PL831) and CLOCK (PL828) were constructed in pBTM116 by an identical approach. Plasmids harboring the full-length ORFs of MOP3, MOP4, and CLOCK were constructed by PCR amplification of the ORF of each cDNA and cloned into the appropriate vectors for expression in yeast or mammalian systems. For yeast expression of full-length proteins, PCR products were cloned into the appropriate sites of pSGBCU11. For mammalian expression, PCR products were cloned into pSVSport and pTarget. The yeast expression vector for full-length ARNT has been described (PL574) (Example 1). The yeast expression vector for full-length MOP3 was designated PL694. Mammalian expression vectors for ARNT (PL87), HIF1α (PL611), and HIF2α (PL447) have been described (Example 1; Jain et al., 1994, J. Biol. Chen 269:31518-31524). Mammalian expression vectors were constructed for MOP3 (PL691 and PL861), MOP4 (PL695 and PL871), and CLOCK (PL941).

Two-Hybrid cDNA Library Screen. The yeast interaction trap was performed using the yeast strain L40 *(MATa, his3Δ200, trp1-901, leu2-3, 112, ade2, LYS::lexAop$_4$HIS, URA3::lexAop$_8$lacZ)* or AMR70 *(MATα, his3, lys2, trp1, leu2, URA::lexAop$_8$-lacZ)* as described (Example 1; Carver and Bradford, 1997, J. Biol. Chem. 272:11452-11456; Vojtek et al., 1993, Cell 74:205-214). The bait plasmid (PL859) was a fusion of the bHLH-PAS domain of

MOP4 to the DNA binding domain of LexA (Hogenesch et al., 1997, *supra*). The MOP4 bait construct was used to screen a human fetal brain cDNA library fused to the transactivation domain of Gal4 (CLONTECH) and transformants were plated on selective media (minus tryptophan, uracil, histidine, and leucine). The cDNAs from surviving colonies, positive for *lacZ* activity were sequenced by the chain termination method (Sanger et al., 1977, PNAS 74:5463-5437). These sequences were analyzed using the BLAST algorithm (Altschul et al., 1990, J. Mol. Biol. 215:403-410).

Interaction Screen Against Known bHLH-PAS Proteins. LexAbHLH-PAS fusion proteins ("baits") of HIF1α, HIF2α, MOP3, MOP4, AHR, ARNT, and CLOCK were transformed into the L40 strain of yeast. The full-length ("fish") MOP3 and ARNT plasmids were transformed into the AMR70 yeast strain, and these transformants were plated onto yeast complete media plates (Kaiser et al., 1994, in *Methods in Yeast Genetics*, Cold Spring Harbor Press, Plainview, NY). The L40 yeast harboring the bait constructs were replica plated onto these yeast complete media plates and mated for 8 hr at 30°C. The plates were then replica plated onto selective media and grown for an additional day at 30°C. 5-Bromo-4-chloro-3-indolyl 13-D-galactoside (X-Gal) overlay assays were performed to determine the relative expression of the *lacZ* reporter gene (Bohen and Yamamoto, 1993, PNAS 90:11424-11428). Western blot analysis, using LexA-specific sera, was performed on extract from each transformant to confirm expression of the fusion protein (see Example 1).

DNA Binding Specificity. To determine high-affinity DNA binding sites for MOP3-MOP4 heterodimers, site selection and amplification was performed as described (Swanson et al., 1995, J. Biol. Chem. 270:26292-26302). Briefly, reticulocyte lysate expressed MOP3 and MOP4 proteins (~0.5 fmol each) were incubated with DNA oligonucleotide randomers corresponding to ~7

$X10^7$ different nucleotide sequences. Randomers were generated and amplified by PCR using oligonucleotides OL990 and OL991 as primers and OL992 as template. After incubating the complexes with the randomers for 30 min at 30°C, samples were loaded directly on 4% polyacrylamide-TBE (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.0) gels to separate MOP3-MOP4 bound DNA from free DNA (Swanson et al., 1995, *supra*). Gel slices corresponding to the migration of bound DNA were excised, incubated overnight in TE (10 mM Tris/1 mM EDTA, pH 8.0), and the eluate subjected to additional PCR using oligonucleotides OL990 and OL991.

Cell Culture and Transient Transfection. Transient transfections of COS-1 cells were performed by the Lipofectamine protocol (GIBCO) as described in Example 1. To mimic hypoxia, 100 μM of cobalt chloride was included in the cell growth media and incubated at 37°C until harvest. To monitor the transcriptional activity of the MOP3-MOP4 or MOP3-CLOCK heterodimers, a synthetic reporter was constructed by annealing phosphorylated oligonucleotides OL1116 and OL1117 and cloning them into the *Bgl*II site in the reporter plasmid pGL3p (Promega). To measure the transcriptional activity of the MOP3-HIF1α or MOP3-HIF2α heterodimers, a synthetic reporter was constructed by annealing phosphorylated oligonucleotides OL1155and OL1156 and then cloning them as above. Luciferase levels were reported in relation to β-galactosidase activity as described in Example 1.

mRNA Expression Analysis. To generate antisense riboprobes, partial cDNAs of the mouse MOP3 and MOP4 were cloned into plasmid vectors harboring bacteriophage promoters. A partial 1.2-kb mouse fragment of MOP3 was obtained by PCR of a mouse kidney cDNA library using oligonucleotides OL579 and OL656, and cloned into pGEM-T in the T7 orientation. For MOP4, reverse transcription-PCR was performed on 3 μg of E17.5d placenta total RNA with oligonucleotides OL522 and OL595.

The resultant fragment was subcloned in pGEM-T in the 5P6 orientation. Total RNA from various mouse tissues was prepared using the Trizol reagent (GIBCO/BRL) according to manufacturer's protocols. Ribonuclease protection assay (RPA) was performed as described for both MOP3 and MOP4 (Luo et al, 1997, Gene Expression 6:287-299). For *in situ* analysis, sense and antisense MOP3 and MOP4 riboprobes were generated with [α-[^{35}S]thio]UTP, 80 μCi (Amersham, >1,000 Ci/mmol; 1 Ci = 37 GBq) as the radioactive ribonucleotide and subjected to alkaline hydrolysis for 13 min at 60°C as described (Jain et al., 1998, Mech. Dev. 73:117-123). Tissue sections (5 μm) were processed and hybridized with the specific riboprobes (Jain et al., 1998, *supra*).

RESULTS

MOP4 Two-Hybrid Library Screen. The MOP4 bait plasmid was used to screen a human fetal brain cDNA library fused to the transactivation domain of Gal4. After screening ~7 x 10^5 colonies, 21 survived selection and were blue in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactoside. BLAST searches revealed that seven of these clones represented four independent MOP3 cDNA fragments. These cDNAs differed in their first 57 codons from the MOP3 cDNA we have described previously (GenBank accession no. U60415; SEQ ID NO:3). These 57 amino acids are identical to that reported by a second group, and appear to be derived from a second promoter (Ikeda and Nomura, 1997, Biochem. Biophys. Res. Commun. 233:258-264). All subsequent functional studies were done using constructs derived from the MOP3 cDNAs identified by the yeast interaction trap.

MOP3 and MOP4 Screened Against Known bHLH-PAS Proteins. To confirm the specificity of the MOP3-MOP4 interaction, we reversed the interaction trap strategy and screened full-length MOP3 against all bHLH-PAS proteins available in this laboratory. As a positive

control we compared these results to a parallel screen using full-length ARNT. Western blot analysis using anti-LexA sera indicated approximately equal expression levels for all fusions. The full-length MOP3 protein interacted strongly with LexAbHLH-PAS fusions of MOP4, CLOCK, and HIF2α and weakly with HIF1α (Figure 5). No interaction of full-length MOP3 could be detected with LexA fusions of MOP3, AHR, ARNT, or the LexA control. Full-length ARNT demonstrated robust interactions with HIF2α and the AHR, and weaker interactions with HIF1α. We did not detect full-length ARNT interactions with LexAbHLH-PAS fusions of MOP3, MOP4, CLOCK, ARNT, or the LexA control (Figure 5).

DNA Binding Specificity of the MOP3-MOP4 Heterodimer. We performed a selection and amplification protocol to identify the DNA sequence bound with high-affinity by the MOP3-MOP4 complex. After three rounds of selection and amplification, a gel shift assay was performed using radiolabeled selected randomers to identify the migration of the complex. We identified a species dependent on the presence of both proteins. A band corresponding to this migration was excised from the polyacrylamide gel, and used as template for a fourth round of amplification before cloning the pool. Analysis of the sequencing data from 10 clones revealed that the MOP3-MOP4 heterodimeric pair bound the sequence G/TGA/GACACGTGACCC (Figure 6). This sequence is an imperfect palindrome containing a core E-box enhancer element (defined as CANNTG, underlined) and specificity for nucleotides in the flanking region (e.g., +4 "A"). We refer to this response element bound by the MOP3-MOP4 as M34. To demonstrate sequence binding specificity and to confirm the selectivity for the +4 nucleotide, we performed competition experiment varying the +4 position to A, C, G, or T (Figure 6). In agreement with our selection results, we observed a strong preference for the flanking +4 "A" nucleotide by the MOP3-MOP4 complex.

MOP3 Forms Transcriptionally Active Complexes with MOP4 and CLOCK. To demonstrate that both MOP3 and MOP4 are required for binding to the M34 element, we performed additional gel shift experiments. A specific band was present only with the combination of MOP3 and MOP4, and was not present with either protein alone. As an additional specificity control, affinity-purified anti-MOP3 or anti-MOP4-specific Igs were used in gel shift experiments. Both MOP3-specific and MOP4-specific IgG were capable of retarding the mobility of the MOP3-MOP4 complex, while purified preimmune IgG alone was not.

To determine whether the MOP3-MOP4 complex could drive transcription *in vivo,* we constructed a vector with three copies of the M34 element upstream of a minimal simian virus 40 promoter-luciferase reporter. Upon cotransfection of the reporter plasmid into COS-1 cells with MOP3 and MOP4, we observed that this combination enhanced transcription 3.3-fold, while neither protein alone was capable of driving transcription over control. The observations that CLOCK also interacted with MOP3 in the yeast interaction trap (Figure 5) and that CLOCK shares extensive homology with MOP4 prompted us to determine if MOP3-CLOCK complex could also drive transcription *in vivo* from an M34 element. Cotransfection of MOP3 and CLOCK revealed that this complex was also active, driving transcription 5.6-fold over control. Transfections with MOP3, MOP4, CLOCK, and ARNT alone, as well as combinations of ARNT and MOP3 or MOP4 failed to drive transcription over control.

MOP3 Forms Functional DNA Binding Complexes with HIF1α and HIF2α. Prompted by our yeast interaction results, we set out to determine the ability of MOP3 to form DNA binding complexes with HIF1α *in vitro.* Because of the asymmetry at the +4 position of the M34 element, we were uncertain which half-site was bound by MOP3. Therefore, we synthesized enhancer elements with the HIF1α 5' half site (TAC) fused to both of the potential

MOP3 3' half-sites described above (GCCCTACGTGACCC or GCCCTACGTGTTCC). We found that the HIF1α/MOP3 complex preferred the GCCCTACGTGACCC element *in vitro*, suggesting that MOP3 preferred an "A" at the +4 position. Therefore the corresponding response element bound by the HIF1α-MOP3 complex, which we refer to as M13, was used in subsequent experiments. The results demonstrate that the M13 element is bound in the presence of the MOP3-HlF1α combination, but not by either protein alone. MOP3-specific and HIF1α-specific antisera abolished this complex while preimmune IgG did not. For comparison we included ARNT in these experiments, and found that ARNT-HIF1α band was more intense than the MOP3-HIF1α complex when all proteins were used at equimolar concentrations.

To determine if MOP3 formed a transcriptionally active complex with either HIF1α or HIF2α *in vivo*, we constructed a synthetic reporter using six copies of the M13 element described above. The M13 reporter was up-regulated when cotransfected with combinations of MOP3-HIF1α and MOP3-HIF2α (3.3-fold and 3.6-fold, respectively). ARNT formed more active complexes with both HIF1α and HIF2α (14.1-fold and 8.1-fold, respectively), consistent with our *in vitro* results. Like ARNT, upon exposure of these transfected cells to cobalt chloride to simulate cellular hypoxia, MOP3 interacted and drove transcription in complexes with both HIF1α and with HIF2α.

Coexpression of MOP3, MOP4, and HIF1α in Neonatal and Adult Murine Tissues. To determine if MOP3 was coexpressed with MOP4 in any murine tissue, ribonuclease protection assays (RPA) and *in situ* hybridization analysis were performed. Parallel RPA analysis of neonatal and adult tissues indicated that MOP3 was most highly expressed in brain, thymus, and muscle. MOP4 showed highest expression in the brain. We performed *in situ* hybridization analysis on tissues where RPA data indicated overlap between MOP3 and MOP4, or MOP3

and HIF1α. Sense controls were negative in all tissues except eye, where the pigment of the retina gave a nonspecific signal. In transverse sections of E15.5 brain, we observed that both MOP3 and MOP4 showed their highest expression levels in the thalamus. In E15.5 eye, we were able to detect colocalization of MOP3 and HIF1α in the retina, but were unable to detect specific labeling of MOP4. The results show that both MOP3 and HIF1α are colocalized in the thymic cortex of postnatal animals. Prompted by the observation of others that the MOP4 mRNA is expressed at low levels in the colon, we assayed that target tissue and observed that MOP4 and HIF1α were coexpressed in postnatal colonic mucosa, while MOP3 was undetectable there (Zhou et al., 1997, PNAS 94:713-718).

DISCUSSION

In an effort to determine the pairing rules of MOP3 and MOP4, we employed the yeast interaction trap to identify the bHLH-PAS partners of these orphans. Our initial experiment using a MOP4 bait construct to screen a brain cDNA library identified MOP3 as a partner. In further experiments, we reversed this approach and used full-length MOP3 to detect interactions with other bHLH-PAS members. This analysis confirmed the MOP3-MOP4 interaction and also demonstrated that CLOCK, HIF1α and HIF2α were additional partners of MOP3. As demonstrated previously, ARNT interacted with the AHR, HIF1α, and HIF2α, but not with MOP4 or CLOCK. The fact that both MOP4 and CLOCK interacted with MOP3 was not surprising given their 75% amino acid sequence identity in their bHLH-PAS domains. The observation that MOP3 was a partner of both HIF1α and HIF2α, but that it did not dimerize with the AHR in the yeast interaction trap was an unexpected result. Due to lack of expression in our yeast system, we were unable to examine the interactions of MOP3 or MOP4 with a number of additional bHLH-PAS

proteins, including mSIM1, mSIM2, hARNT2, and hSRC1. Thus, we do not exclude the possibility that additional MOP3 and MOP4 interactions with these proteins may be important. Nevertheless, our data lead us to suggest that MOP3 is a general partner of a number of bHLH-PAS factors, with a distinct interaction profile from that of the more well characterized general partner ARNT.

Analysis of MOP3 and MOP4 revealed that these proteins did not share perfect identity with any other known bHLH proteins in their basic residues thought to contact DNA. Therefore, we could not readily predict the response elements that the MOP3-MOP4 heterodimer would bind. To overcome this limitation, we employed a DNA selection and amplification protocol and determined that the MOP3-MOP4 complex bound an E-box, with flanking region specificity for an "A" at +4 (i.e., CACGTGA, M34 element). In keeping with our prediction that MOP4 and CLOCK are functional homologs, transfection experiments demonstrated that the combination of either MOP3-MOP4 or MOP3-CLOCK was capable of driving transcription from M34 elements, while neither MOP3, MOP4, or CLOCK alone displayed this activity. In support of our argument that MOP3 harbors a partnering specificity distinct from that of ARNT, we observed that neither MOP3 nor MOP4 was capable of interacting with ARNT and driving transcription from the M34 element in its presence.

What could be the consequence of these interactions? Experiments from a number of laboratories indicate that circadian behavior may be regulated at the transcriptional level by complex interactions between multiple PAS domain containing proteins. Strong genetic evidence supports a role for CLOCK in the maintenance of circadian behavior in mice and the product of the *period* gene (PER) for control of circadian rhythms in *Drosophila*. The fact that MOP4 is a brain specific homolog of CLOCK and that these factors share MOP3 as a common dimeric partner suggests that both MOP3 and MOP4

may play a role in this process as well. In addition to the mammalian MOP3, MOP4 and CLOCK proteins, murine and human homologs of *Drosophila* PER have recently been characterized. Like *Drosophila* PER, the mRNA levels of these mammalian homologs respond to light and display circadian rhythmicity. Sequence analysis of PER proteins indicates that they contain PAS domains, but do not contain consensus bHLH domains. Coupled with additional biochemical evidence from others, these data suggest that PER proteins may affect their own transcription through interactions mediated by their PAS domains. Thus, these PERs may impact transcriptional activity of other bHLH-PAS dimers by acting as either dominant negative inhibitors that block pairing of trancriptionally active complexes, or they may act in a positive manner as coactivators of these complexes.

In addition to defining the pairing rules and DNA binding specificities of MOP3 and MOP4, our data lead us to a testable model that describes circadian oscillation of transcription. Without intending to limit the present invention by any particular explanation of mechanism, we speculate that MOP3-CLOCK or MOP3-MOP4 complexes regulate PER transcription through CACGTGA-containing enhancers. The transcriptional activity of these promoters could in turn be modified by feedback inhibition/activation by the PER protein products themselves. In support of this idea, an E-box element in the Drosophila PER promoter, required for normal cycling of the PER mRNA, bears striking resemblance to the M34 element we have identified (i.e., 5'-CACGTGAGC-3' compared with 5'-CACGTGACC-3'). Given that we are borrowing from both *Drosophila* and mammalian systems, our model assumes that these signal transduction pathways have been largely conserved throughout evolution. In keeping with this idea, a search of *Drosophila* expressed sequence tags revealed the existence of an uncharacterized MOP4/CLOCK homolog (GenBank accession no.

AA698290) and an uncharacterized MOP3 homolog (GenBank accession no. AA695336).

What could be the consequences of MOP3-HIF interactions? Transient transfection experiments showed that MOP3 formed transcriptionally active complexes with HIF1α and HIF2α and that these complexes responded to cellular hypoxia. MOP3 may play a specialized role in hypoxia signaling. The different tissue specific expression profiles of MOP3 and ARNT suggests that MOP3 may regulate cellular responses to hypoxia at distinct sites, such as the retina, thymic cortex, and thalamus. Moreover, the observation that MOP3 binds a GTG half-site with flanking region specificity for an "A" at +4, may indicate that MOP3/HIF complexes may have greater affinity for a distinct subset of hypoxia response elements (i.e., TACGTGA vs. the more commonly observed TACGTGG elements observed in known hypoxia responsive enhancers). Finally, the observation that MOP4 is expressed at a site where MOP3 expression appears low, i.e., colonic mucosa, suggests that additional partners may exist for MOP4 and CLOCK and that all bHLH-PAS signaling pathways may involve complex equilibria between multiple PAS proteins.

EXAMPLE 3
Chromosomal Localization and Molecular Characterization of MOP7 as a Third Hypoxia Inducible Factor

Hypoxia inducible factors (HIFs) regulate transcriptional responses to low oxygen tension and other physiological conditions that rely upon glucose for cellular ATP. The HIFs are heterodimeric transcription factors that are composed of two bHLH-PAS proteins. The bHLH-PAS subunits can be classified as α-class or β-class. In addition to amino acid sequence similarity, the most distinguishing characteristic of the α-class subunits is that they are rapidly up-regulated by cellular hypoxia, or treatment with iron chelators and

certain divalent cations (e.g. Co^{++}). The previously described α-class subunits are referred to as HIF1α (MOP1 herein) and HIF2α (MOP2 herein). In contrast, the β-subunits appear to be constitutively expressed and ready to pair with their up-regulated α-class partner. Recent evidence suggests that ARNT, ARNT2 and MOP3 are prototype β-class subunits. At the present time, a number of well-characterized HIF-responsive gene products have been identified. These genes include those encoding EPO, VEGF and GLUT1, among others. The promoters of these genes are regulated by HRE elements that are recognized by the HIFαβ heterodimer. The HREs contain the core TACGTG element and are found both 5' and 3' to the regulated promoter in a number of hypoxia responsive genes.

It is of academic and practical interest to understand how bHLH-PAS proteins signal, as well as the biological consequences that result from the sharing of bHLH-PAS partners. The recent generation of thousands of expressed sequence tags (ESTs) has provided the opportunity to identify and classify orphan HIF subunits based upon nucleotide sequence similarity with known bHLH-PAS proteins. As the result of these efforts, we have identified, and describe the cloning and characterization below, of a third HIF α-class subunit, referred to above as MOP7. For consistency of nomenclature, this protein also is referred to as "HIF3α". Using a number of biochemical approaches, we demonstrate that the MOP 7 (HIF3α) cDNA encodes a protein that meets the major criteria of an α-class HIF subunit. The observation that multiple HIF α and β subunits are encoded by the mammalian genome suggests that a complex array of subunit interactions and tightly controlled developmental expression patterns governs transcriptional response to hypoxic stress.

MATERIAL AND METHODS

Gel-shift oligonucleotides. The complementary oligonucleotide pairs used in gel-shift assays are shown below (5' to 3'). They contain a constant flanking sequence and the wildtype or mutant HRE core sequence (underlined):

OL396 TCGAGCTGGGCAGGTAAGGTGGCAAGGC
OL397 TCGAGCCTTGCCACGTTACCTGCCCAGC
OL398 TCGAGCTGGGCAGGTGAGGTGGCAAGGC
Ol399 TCGAGCCTTGCCACGTCACCTGCCCAGC
OL414 TCGAGCTGGGCAGGGTAGGTGGCAAGGC
OL415 TCGAGCCTTGCCACGTACCCTGCCCAGC

PCR Oligonucletides:

OL1014	GCCATGGCGTTGGGGTGCAG
OL1017	ACTGTGTCCAATGAGCTCCAG
OL1178	GCCTCCATCATGCGCCTCACAATCAGC
OL1210	CCCCGTTACTGCCTGGCCCTTGCTCA
OL1323	AGCCGAGGGGGTCTGCGAGTATGTTGC
OL1324	GCTGCTGACCCTCGCCGTTTCTGTAGT
OL1397	GTCGACGCCACCATGGACTGGGACCAAGACAGG
OL1427	GGATCCTCAGTGGGTCTGGCCCAAGCC
OL1548	GCGGGGTGCTGGGAGTGGCTGCTAC
OL1698	GCCTTCCTGCACCCGCCTTCCCTGAG
OL1769	GCGGCCGCAAAAAACAAGACCGTGGAGACA
OL1771	GCCCTGGGAGAATAGCTGTTGGACTTTGGGCAATTGCTCACT
OL1772	GCGGCCGCCTATTCTGAAAAGGGGGGAAA
AP1	CCATCCTAATACGACTCACTATAGGGC
AP2	ACTCACTATAGGGCTCGAGCGGC

Cloning of MOP7. TBLASTN and BLASTX algorithms were used to search nucleotide sequences corresponding to amino acids 54 to 125 of hHIF1α (http://dot.imgen.bcm.tmc.edu: 9331/seq-search/Options/blast.html) (Hwang et al., J. Mol. Cell. Cardiol. 26: 1329-1333, 1994). One mouse EST clone, Genbank Accession AA028416 (designated PL773), was found to encode a novel bHLH-PAS protein. To obtain the complete open reading frame, we performed a series of PCR amplifications using primer-anchored cDNA derived from mouse lung ("Marathon-Ready," Clontech) (Siebert et al., Nuc. Acids Res. 23: 1087-1088, 1995). A 3' rapid amplification of cDNA ends (RACE) reaction was performed using oligonucleotides OL1178 and anchor primer AP1. The product of this reaction was reamplified in a second reaction with OL1178 and AP2. The 2.0-kb 3' PCR product obtained by this protocol was cloned into the

pGEM Teasy vector (Promega) and designated PL970. The clone was sequenced and found to contain an ORF followed by a translational stop site (Figure 7). To confirm the position of the traslational stop site, OL1324 was used in an independent 3' RACE reaction. The 0.9 kb product was cloned into a pGEM Teasy vector (PL1017) and was found to contain the same stop codon (Figure 7). To obtain the 5' end of the cDNA, OL1323 was used in a RACE reaction against oligonucleotide AP1. The 1.2 kb RACE product was cloned into pGEM Teasy vector (PL1016) and found to contain a translation start codon ATG followed by a long ORF. We define the nucleotide A from the initiation codon as position 1 of the cDNA. In addition, the translational start site is defined by the presence of an in-frame stop codon 51 nucleotides upstream. To generate expression plasmids containing the full ORF, a PCR reaction was performed using OL1210 and OL1397 with PL1016 as template. The PCR fragment was cloned into pGEM Teasy vector in the SP6 orientation and named PL1024. The *Nde*I digested PL1024 was then inserted into the *Nde*I digested PL970 to generate the full-length HIF 3α in the pGEM-Teasy vector (PL1025).

Construction of MOP7 expression plasmids. For MOP7 expression in mammalian cells, the ORF was amplified by PCR using OL1397 and OL1427 with PL1025 as template. The resultant plasmid was cloned into pTarget vector downstream of the CMV promoter (Promega) and was named PL1026 (Figure 7).

To confirm the hypoxia inducibility of MOP7, we constructed a fusion protein comprised of the DNA binding domain from GAL4 (residues 1-147), the predicted hypoxia responsive domain-1 (HRD1) from mMOP7 (residues 453-496), and the transactivation domain (TAD) from hARNT (residues 581-789). The HRD1 was amplified using OL1769 and OL1771 with mMOP7 as template. To form the HRD1/TAD chimera, the resultant PCR fragment from above was used as a megaprimer in a second PCR reaction with OL1772 as the

second primer and hARNT as the template (Barik et al., Biotechniques 10: 489-490, 1991). The HRD1/TAD chimeric fragment was cloned into the *Not*I site of the GAL4 fusion vector pBIND (Promega) and designated PL1131.

Structural gene analysis and chromosomal localization. The MOP7 insert from PL773 was cut with *Eco*RI/*Nde*I and the 0.6-kb fragment was purified and used as a probe to screen for bacterial artificial chromosome (BAC) clones containing the mouse MOP7 gene (Genome Systems, Inc.). Oligonucleotides derived from the mMOP7 sequence were used as primers to sequence the BAC DNA, and the splice sites were deduced by comparing the genomic and cDNA sequences. To obtain BACs containing the human MOP7, oligonucleotides OL1014 and OL1017 were used in a PCR reaction with human heart cDNA as template (Clontech) to amplify a MOP7 fragment (Genbank Accession No. AF079154). This fragment was subcloned into the pGEM-Teasy vector, confirmed by sequencing, and used as a probe to screen for a BAC clone harboring the human structural gene for MOP7 as above. The identity of the resultant BAC was confirmed by direct sequencing using primers specific for hMOP7. The human MOP7 chromosomal location was identified by PCR reactions against human/hamster somatic cell hybrid DNA using human MOP7-specific oligonucleotides. This location was confirmed by fluorescence *in situ* hybridization (FISH) using the BAC harboring human MOP7 structural gene as the probe (Genome Systems, Inc.).

Northern Blot analysis. To generate a hybridization probe for northern blot analysis, the 894 bp MOP7 insert from PL1017 was excised with *Eco*R1 and radiolabeled with [α-^{32}P]dCTP by random priming. A northern blot containing poly A$^+$ mRNA from different mouse tissues (Origene Technologies, Rockville, MD) was hybridized with 5 x 10^6 cpm/ml MOP7 probe. β-actin was used as a loading control.

Gel-shift assay. To generate a double strand oligonucleotide probe containing the core HRE element, 50 ng of oligonucleotide OL414 was end-labeled with [γ-^{32}P]ATP and was annealed with 10 fold excess of cold complementary oligonucleotide OL415. Unlabeled oligonucleotides containing either wild-type HRE sequence (TACGTG) or mutant HRE sequences, AACGTG (OL396/397) or GACGTG (OL398/399), were used in competition experiments to demonstrate specificity. For expression of the bHLH-PAS proteins, mMOP7 (PL1025) and hARNT (PL87) were synthesized in a reticulocyte lysate in the presence of [^{35}S]methionine. The amount of each protein synthesized was calculated by measuring radioactivity and estimated to be approximately 1 fmol in each 10 μl gel-shift reaction. Each gel-shift assay was performed with 100,000 cpm of oligonucleotide probe per 10 μl reaction. To confirm complex identity, 1 μl of anti-ARNT sera was used to supershift the DNA bound protein complex.

Cell culture and transfection. COS-1 cells were maintained in high glucose DMEM medium supplemented with 10% fetal calf serum, 100 units/ml penicillin and streptomycin. The HRE driven luciferase reporter (PL945) was made by annealing OL1174 and OL1175 and then cloning the fragment into pGL3-promoter vector (Promega, Madison, WI). For transient transfection experiments, mammalian expression plasmids expressing MOP7 or hARNT with the reporter using lipofectamine (GIBCO BRL Life Technologies). A β-galactosidase-expression plasmid was co-transfected to control for transfection efficiency. Cells were incubated for 20-24 hours with or without treatment of cobalt chloride or hypoxia (1% O_2) before being harvested. The luciferase and β-galactosidase activities were determined using the luciferase assay and the Galacto-Light protocols as previously described (see Example 2).

RESULTS and DISCUSSION

From a TBLASTN search of the dBEST database with the sequence corresponding to amino acid residues 54 to 125 of hHIF1α, we identified a mouse EST clone (AA028416) that appeared to be a novel bHLH-PAS protein. This protein is referred to herein as MOP7 and as HIF3α, to denote its relationship to other hypoxia-inducible factors. To obtain the complete ORF of this cDNA, a series of RACE reactions was performed using cDNA from mouse lung as template. The MOP7 ORF (SEQ ID NO:7) spans 1.98 kb and encodes a 662-amino acid protein (SEQ ID NO:16) with a predicted molecular weight of 73 kDa. Northern blot analysis on mRNA prepared from selected mouse tissues identified a MOP7 transcript that is expressed in adult thymus, lung, brain, heart and kidney. This expression pattern is distinct from that reported for other α-class HIFs. HIF1α is most abundantly expressed in kidney and heart, and HIF2α is most abundantly expressed in vascular endothelial cells and is highest in lung, placenta and heart.

HIF1α (MOP1) is the most well-characterized α-class subunit. This protein can be described by a number of signature motifs. In addition to the bHLH-PAS domains, HIF1α also contains two HRD motifs in its C-terminus that confer hypoxia responsiveness. The HRD1 appears to primarily confer hypoxia-dependent protein stability, whereas HRD2 appears to confer hypoxia-responsive transcriptional activity. To determine if similar motifs occur in MOP7, we compared HIF1α, HIF2α and MOP7 protein sequences using the CLUSTAL algorithm (Higgins & Sharp, Gene 73: 237-244, 1988. These three sequences shared greater than 92% identity in the basic region, 68% in the HLH domain, and greater than 53% in the PAS domain. Although little sequence with significant homology to HRD2 was found, a 36-amino acid stretch within the C-terminal half of MOP7 was found to share 61% identity with the HRD1 of HIF1α.

To further demonstrate the evolutionary relationship between these α-class HIFs, we compared their gene structure and chromosomal localization. Direct sequencing of a BAC clone containing the mMOP7 gene revealed 15 exons, all with consensus splice donor/acceptor sites (see sequences of Genbank Accession No. AF079140-079153 for exon-intron junctions). We found that 11 of 15 and 10 of 15 splice junctions found in the mMOP7 gene are conserved to those found in hte structural genes of mHIF1α and hHIF2α, respectively (Figure 8). To characterize the distribution of HIF genes in the mammalian genome, we used human MOP7-specific PCR reactions against a human/hamster somatic cell panel and mapped the MOP7 gene locus on human chromosome 19. This locus was further defined to chromosome 19q13.13-13.2 by FISH using a BAC clone containing the human MOP7 structural gene as a probe. Therefore, the human MOP7 locus is distinct from that of human HIF1α and HIF2α, which reside on chromosome 14q21-24 and 2p16-21, respectively.

As a biochemical proof that MOP7 was a bona fide α-class HIF, we performed gel-shift and transient transfection analyses. Because HIF1α and HIF2α are known to pair with the β-class HIF subunit ARNT, we predicted that MOP7 would also pair with ARNT. Based upon sequence identity in their basic regions, we also predicted that a MOP7-ARNT would bind the HRE core sequence, TACGTG. As predicted, the gel-shift analysis showed that MOP7 only bound to the HRE containing oligonucleotide in the presence of ARNT. The specificity of the interaction was demonstrated by two additional observations. First, the MOP7-ARNT-HRE complex was abolished by anti-ARNT IgGs but not by preimmune antibodies. Second, the complex was blocked by an excess of HRE containing oligonucleotide but not by oligonucleotides with a single mutation within the core HRE sequence. To determine if this interaction could also occur *in vivo*, MOP7 and/or ARNT were

cotransfected into COS-1 cells with a luciferase reporter driven by six HRE enhancer elements. The results demonstrated that the combination of MOP7 and ARNT upregulated transcription from the HRE-driven reporter by 11.7 fold, whereas neither protein alone had an effect. In addition, the activity of these complexes was enhanced by either hypoxia or cobalt chloride.

To demonstrate that the MOP7 activity was directly upregulated by hypoxia, we employed a fusion protein approach that has been used to map the HRDs of HIF1α. HRD1 of HIF1α has been shown to encode a hypoxia-responsive protein stability domain that also displays weak transcriptional activity. Given the sequence similarity between residues 453-496 if MOP7 and the HRD1 of HIF1α, we predicted that this domain would independently respond to hypoxic stimulus or cobalt ion exposure. To test this, we constructed a plasmid expressing a fusion protein comprised of the DNA binding domain of GAL4, the predicted HRD1 of MOP7, and the TAD from ARNT. We predicted that we could measure the response of this domain by monitoring the output from a GAL4-driven luciferase reporter in Hep3B cells. The results demonstrated that the fusion protein's activity increased by 4.5- and 4.2-fold, upon treatment with cobalt chloride or hypoxia, respectively. In control experiments, we observed that a GAL4 fusion protein harboring only the ARNT-TAD did not respond to either hypoxia or cobalt chloride treatment. The level of inducibility seen with the HRD1 fusion is consistent with that obtained for a similar fusion protein using the HRD1 domain of HIF1α. This result provided evidence that amino acids 453 to 496 of MOP7 was sufficient to confer the hypoxia inducibility and that the stability of the parent protein is regulated in a manner that is similar to that of HIF1α and HIF2α.

In eucaryotes, transcriptional responses to low oxygen tension are mediated by complex interactions

between a number of α- and β-class HIF subunits. The characterization of a third α-class HIF with a tissue distribution that is distinct from either HIF1α or HIF2α provides evidence that cellular responses to hypoxia result from a complex set of interactions from multiple combinations of αβ pairs. MOP7 also may have a distinct role in mediating biological responses to hypoxia. In support of this notion, MOP7 and HIF1α have limited sequence homology in their C-termini. Most importantly, MOP7 contains sequence that corresponds to HIF1α's protein stability element, HRD1, but not to its hypoxia-responsive TAD element, HRD2. This may indicate that MOP7-ARNT complexes have decreased transcriptional potency relative to other HIF heterodimers. The importance of this complexity is underscored by the presence of HIF1α, HIF2α and MOP7 in both mice and humans. Finally, this complexity appears to be highly conserved among vertebrates. In support of this idea, we have cloned a partial human MOP7 cDNA and have shown all three HIF α-class genes reside on separate human chromosomes and display considerable sequence divergence in their C-termini.

The present invention is not limited to the embodiments described and exemplified above, but is capable of variation and modification within the scope of the appended claims.